PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K	A2	(11) International Publication Number: WO 00/39149 (43) International Publication Date: 6 July 2000 (06.07.00)

(21) International Application Number: PCT/US99/31157

(22) International Filing Date: 29 December 1999 (29.12.99)

(30) Priority Data:
09/224,246 30 December 1998 (30.12.98) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor: BARNES, Thomas, M.; 22 Hanson Street #2, Boston, MA 02118 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

Ref "AJ" DX0903K - J. Fernando Bazan
09/399,492

(54) Title: SECRETED PROTEINS AND USES THEREOF

```
                                                                             V
GTCGACCCACGCGTCCGGGGAGCTGGTCCCTGCGCTCCCTGCGCCCTGGGGATGCCCCTGCCGCCCTGACGCCCGCCAG   79

CCTGAGCCACCGGCGCATGTGACCGCGCGTCCGCCCCAGTCCCATCCGTAGGCGCCCGGCGCCCGGCCCCGCAGCGGCC   158

                                                                M   Q   A   R       4
TCGTTGTCCCCGCCGGCCCCCGCCCGGTCTCCCGCGCTGCCACCCGCCGCCGGCCCTGCCGCC ATG CAG GCG CGA   233

 A   L   L   L   A   A   L   A   A   L   A   L   A   R   E   P   P   A   A   P     24
GCG CTG CTC CTG GCC GCG TTG GCC GCG CTG GCG CTG GCC CGG GAG CCC CCT GCG GCG CCG    293

 C   P   A   R   C   D   V   S   R   C   P   S   P   R   C   P   G   G   Y   V     44
TGT CCC GCG CGC TGC GAC GTG TCG CGG TGT CCC AGC CCC CGC TGC CCC GGC GGC TAC GTG    353

 P   D   L   C   N   C   C   L   V   C   A   A   S   E   G   E   P   C   G   G     64
CCC GAC CTC TGC AAC TGC TGC CTG GTG TGC GCC GCC AGC GAG GGC GAG CCC TGT GGC GGC    413

 P   L   D   S   P   C   G   E   S   L   E   C   V   R   G   L   C   R   C   R     84
CCT CTG GAC TCG CCT TGC GGC GAG AGC CTG GAG TGC GTG CGC GGC CTA TGC CGC TGC CGC    473

 W   S   H   A   V   C   G   T   D   G   H   T   Y   A   N   V   C   A   L   Q    104
TGG TCG CAC GCC GTG TGT GGC ACC GAC GGG CAC ACC TAT GCC AAC GTG TGC GCG CTG CAG    533

 A   A   S   R   R   A   L   Q   L   S   G   T   P   V   R   Q   L   Q   K   G    124
GCG GCC AGC CGC CGC GCG CTG CAG CTC TCC GGG ACG CCC GTG CGC CAG CTG CAG AAG GGC    593

 A   C   P   L   G   L   H   Q   L   S   S   P   R   Y   K   F   N   F   I   A    144
GCC TGC CCG TTG GGT CTC CAC CAG CTG AGC AGC CCG CGC TAC AAG TTC AAC TTC ATT GCT    653
```

(57) Abstract

The invention provides isolated nucleic acid molecules, designated HtrA-2 (high temperature requirement-2) nucleic acid molecules which encode wholly secreted proteins with homology to the human HtrA (high temperature requirement) protein. The invention also provides isolated nucleic acid molecules, designated TANGO 219 nucleic acid molecules which encode wholly secreted proteins. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

SECRETED PROTEINS AND USES THEREOF

Cross Reference to Related Applications

This application is a continuation-in-part of co-pending Application No. 09/224,246 filed December 30, 1998, which is incorporated herein by reference in its entirety.

Background of the Invention

Many secreted proteins, for example, cytokines and cytokine receptors, play a vital role in the regulation of cell growth, cell differentiation, and a variety of specific cellular responses. A number of medically useful proteins, including erythropoietin, granulocyte-macrophage colony stimulating factor, human growth hormone, and various interleukins, are secreted proteins. Thus, an important goal in the design and development of new therapies is the identification and characterization of secreted and transmembrane proteins and the genes which encode them.

Many secreted proteins are receptors which bind a ligand and transduce an intracellular signal, leading to a variety of cellular responses. The identification and characterization of such a receptor enables one to identify both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, permitting one to identify or design modulators of receptor activity, e.g., receptor agonists or antagonists and modulators of signal transduction.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules which encode the TANGO 214 and 219 proteins, both of which are secreted proteins.

The TANGO 214 proteins share significant homology to the human HtrA protein, the human homologue of the *E. coli* HtrA (high temperature requirement) gene product, a critical component of the bacterial response to stress. Because of their homology to the human HtrA protein, TANGO 214 proteins (and the nucleic acids that encode them) are referred to herein as HtrA-2 proteins (and nucleic acid molecules).

The HtrA-2 and TANGO 219 proteins, fragments, derivatives, and variants thereof are collectively referred to as "polypeptides of the invention" or "proteins of the invention.". Nucleic acid molecules encoding the polypeptides or proteins of the invention are collectively referred to as "nucleic acids of the invention."

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this

invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable for use as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the nucleotide sequence of the cDNA insert of a clone deposited with ATCC as Accession Number 98899, or a complement thereof, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention features nucleic acid molecules of at least 510, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, or 2570 nucleotides of the nucleotide sequence of SEQ ID NO:1, the nucleotide sequence of a human HtrA-2 cDNA, the nucleotide sequence of the human HtrA-2 cDNA clone of ATCC Accession No. 98899, or a complement thereof. The invention also features nucleic acid molecules comprising at least 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, or 910 nucleotides of nucleic acids 1 to 925 of SEQ ID NO:1, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, or 1350 nucleotides of the nucleotide sequence of SEQ ID NO:2, or a complement thereof.

The invention features nucleic acid molecules of at least 380, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1575, or 1595 nucleotides of the nucleotide sequence of SEQ ID NO:18, the nucleotide sequence of a mouse HtrA-2 cDNA, or a complement thereof. The invention also features nucleic acid molecules comprising at least 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, or 280 nucleotides of nucleic acids 1 to 285 of SEQ ID NO:18, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 380, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1025, or 1035 nucleotides of the nucleotide sequence of SEQ ID NO:19, or a complement thereof.

The invention features nucleic acid molecules of at least 280, 300, 350, 400, 450,

500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, or 1240 nucleotides of the nucleotide sequence of SEQ ID NO:13, the nucleotide sequence of a human TANGO 219 cDNA, the nucleotide sequence of the human TANGO 219 cDNA clone of ATCC Accession No. 98899, or a complement thereof. The invention also features nucleic acid molecules comprising at least 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides of nucleic acids 1 to 2227 of SEQ ID NO:13, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, or 440 nucleotides of the nucleotide sequence of SEQ ID NO:14, or a complement thereof.

The invention features nucleic acid molecules of at least 300, 305, 310, 315, 320, 325, 330, 335, 340, 345, 350, 355, 360, 365, 370, 375, 380, 385, or 390 nucleotides of the nucleotide sequence of SEQ ID NO:21, the nucleotide sequence of a mouse TANGO 219 cDNA, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least or 300, 305, 310, 315, 320, 325, 330, 335, 340, 345, 350, 355, 360 nucleotides of the nucleotide sequence of SEQ ID NO:22, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200) nucleotides of the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, or 22, the nucleotide sequence of a human HtrA-2 cDNA, the nucleotide sequence of a human TANGO 219 cDNA, the nucleotide sequence of the human HtrA-2 cDNA of ATCC Accession Number 98899, the nucleotide sequence of the human TANGO 219 cDNA of ATCC Accession Number 98899, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 50 (60, 80, 100, 130, 150, 180, 200, 230, 250, 280, 300, 350, 380, 400, 420, or 450) nucleotides of the nucleotide sequence of SEQ ID NO:13, the nucleotide sequence of a human TANGO 219 cDNA, the nucleotide sequence of the human HtrA-2 cDNA of ATCC Accession Number 98899, the nucleotide sequence of the human TANGO 219 cDNA of ATCC Accession Number 98899, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NO:3, 15, 20, or 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, the amino acid sequence encoded by the human TANGO 219

cDNA of ATCC Accession Number 988899, or a complement thereof, wherein the protein encoded by the nucleotide sequence also exhibits at least one structural and/or functional feature of a polypeptide of the invention.

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the nucleotide sequence of the human HtrA-2 cDNA of ATCC Accession Number 98899, or the nucleotide sequence of the human TANGO 219 cDNA of ATCC Accession Number 98899.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:3, 15, 20, 23 or a fragment including at least 15 (25, 30, 50, 100, 150, 300, 400, or 450) contiguous amino acids of SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, wherein the fragment exhibits at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:3, 15, 20, 23, or a fragment including at least 15 (25, 30, 50, 70, 90, 100, 120, 140) contiguous amino acids of SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, wherein the fragment exhibits at least one structural and/or functional feature of a polypeptide of the invention.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of a nucleic acid sequence encoding SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, or a complement thereof under stringent conditions, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 60%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by

the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention

Also within the invention are isolated polypeptides or proteins which preferably are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 60%, preferably 65%, 75%, 85%, or 95% identical the nucleic acid sequence encoding SEQ ID NO:3, 15, 20, or 23, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, a complement thereof, the non-coding strand of the nucleotide sequence of the human HtrA-2 cDNA of ATCC Accession Number 98899, the non-coding strand of the nucleotide sequence of the human TANGO 219 cDNA of ATCC Accession Number 98899, wherein such nucleic acid molecules encode polypeptides or proteins that preferably also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID NO:3, 15, 20, 23, the amino acid sequence encoded by the human HtrA-2 cDNA of ATCC Accession Number 988899, or the amino acid sequence encoded by the human TANGO 219 cDNA of ATCC Accession Number 988899, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 or a complement thereof under stringent conditions, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA of human HtrA-2 cDNA of ATCC Accession Number 988899, the cDNA of human TANGO 219 cDNA of ATCC Accession Number 988899, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA of human HtrA-2 cDNA of ATCC Accession Number 988899, the cDNA of human TANGO 219 cDNA of ATCC Accession Number 988899, or a complement thereof, wherein such

nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA of human HtrA-2 cDNA of ATCC Accession Number 988899, the cDNA of human TANGO 219 cDNA of ATCC Accession Number 988899, or a complement thereof, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention. In other embodiments, the nucleic acid molecules are at least 50 (60, 80, 100, 130, 150, 180, 200, 230, 250, 280, 300, 350, 380, 400, 420, or 450) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA of human HtrA-2 cDNA of ATCC Accession Number 988899, the cDNA of human TANGO 219 cDNA of ATCC Accession Number 988899, or a complement thereof, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

In one embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, e.g., recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides host cells containing such a vector or a nucleic acid molecule of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a polypeptide is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, or a functional activity of a polypeptide or nucleic acid of the invention refers to an activity exerted by a protein, polypeptide or nucleic acid molecule of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein, or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein.

For HtrA-2, biological activities include, e.g., (1) the ability to modulate growth factor function, e.g., that of insulin-like growth factor (IGF), e.g., IGF-I, IGF-II, , by, for

example, modulating the availability of growth factors and/or their receptors; (2) the ability to modulate (e.g., inhibit) the activity of a proteolytic enzyme, e.g., a serine protease; and (3) the ability to modulate the function, migration, proliferation (e.g., suppress cell growth), and/or differentiation of cells, e.g., cells in tissues in which it is expressed (see description of expression data below) and, in particular, bone cells such as osteoblasts and osteoclasts, and cartilage cells such as chondrocytes.

Other activities of HtrA-2 include: (1) the ability to act as a proteolytic enzyme cleaving either itself (e.g., autocatalysis, e.g., autocatalysis between its own Kazal and serine protease domains) or other substrates; (2) the ability to bind to an inhibitor of proteolytic enzyme activity, e.g., an inhibitor of a serine protease, e.g., α_1-antitrypsin; (3) the ability to modulate the activity of proteins (e.g., TGF-beta family members) in the activin/inhibin growth factor system; and (4) the ability to perform one or more of the functions of human HtrA described, for example, in Hu et al. (1998) *J. Biol. Chem.* 273(51):34406-34412, the contents of which are incorporated herein by reference.

Other activities of HtrA-2 include: (1) the ability to modulate the function of a normal or mutated presenilin protein (e.g., presenilin-1 (PS-1) or presenilin-2 (PS-2)); and (2) the ability to perform a function of the human serine protease PSP-1, described in EP 828 003, the contents of which are incorporated herein by reference.

Still other activities of HtrA-2 include: (1) the ability to mediate protein-protein interactions; (2) the ability to modulate protein degradation, e.g., degradation of denatured and/or misfolded proteins; (3) the ability to act as a chaperone protein, e.g., to renature misfolded proteins and help to restore their function; (4) the ability to interact with (e.g., bind to) the normal or mutated gene product of a human presenilin gene (e.g., human presenilin 1 (PS-1), e.g., mutant PS-1 TM16TM2 loop domain as described in PCT Number WO98/01549, published January 15, 1998); (5) the ability to interact with (e.g., bind to) a protein expressed in brain; (6) the ability to modulate a neurological function; (7) the ability to interact with (e.g., bind to) a protein containing the following consensus sequence: Xaa-Ser/Thr-Xaa-Val-COO-, where Xaa is any amino acid, Ser is Serine, Thr is Threonine, Val is Valine (which can be substituted with other hydrophobic residues), and COO- is the protein C terminus; and (8) the ability to modulate production and secretion of prostaglandins.

For TANGO 219, biological activities include, the ability to modulate the normal function, migration, proliferation, and/or differentiation of cells in which it is expressed (see description of expression data below).

In one embodiment, a polypeptide of the invention has an amino acid sequence

sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 60% identity, preferably 65% identity, more preferably 75%, 85%, 95%, 98% or more identity are defined herein as sufficiently identical.

In one embodiment, an HtrA-2 protein includes at least one of the following domains: a signal sequence, an IGF binding domain, a Kazal serine protease inhibitor domain, a serine protease domain, and a PDZ domain. In another embodiment, an HtrA-2 protein includes two or more of the following domains: a signal sequence, an IGF binding domain, a Kazal serine protease inhibitor domain, a serine protease domain, and a PDZ domain, and is secreted.

In one embodiment, a TANGO 219 protein includes a signal sequence.

In another embodiment, a nucleic acid molecule of the present invention encodes an HtrA-2 protein which includes at least one of the following domains: a signal sequence, an IGF binding domain, a Kazal serine protease inhibitor domain, a serine protease domain, and a PDZ domain. In another embodiment, a nucleic acid molecule of the present invention encodes an HtrA-2 protein which includes two or more of the following domains: a signal sequence, an IGF binding domain, a Kazal serine protease inhibitor domain, a serine protease domain, and a PDZ domain. Still other embodiments feature a nucleic acid molecule of the present invention which encodes an HtrA-2 protein which includes three, four, or all of the following domains: a signal sequence, an IGF binding domain, a Kazal serine protease inhibitor domain, a serine protease domain, and a PDZ domain.

In another embodiment, a nucleic acid molecule of the present invention encodes a TANGO 219 protein which includes a signal sequence.

The polypeptides of the present invention, or biologically active portions thereof, or antibodies of the invention, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibodies that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies. In addition, the polypeptides of the invention or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small organic molecule.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of the invention wherein a wild-type form of the gene encodes a protein having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression

of the polypeptide or nucleic acid in the presence and absence of the compound.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1A, 1B, and 1C depict the cDNA sequence of human HtrA-2 (SEQ ID NO:1) and the predicted amino acid sequence of HtrA-2 (SEQ ID NO:3). The open reading frame of SEQ ID NO:1 extends from nucleotide 222 to nucleotide 1580 of SEQ ID NO:1 (SEQ ID NO:2).

Figure 2 depicts a hydropathy plot of human HtrA-2. Relatively hydrophobic regions of the protein are above the dashed horizontal line, and relatively hydrophilic regions of the protein are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the amino acid sequence of HtrA-2 is depicted.

Figures 3A-3G depict an alignment of the nucleotide sequence of human HtrA (SEQ ID NO:10; GenBank Accession Number Y07921) and the nucleotide sequence of human HtrA-2 (SEQ ID NO:1). The nucleotide sequences of human HtrA and human HtrA-2 are 50.9% identical. This alignment was performed using the ALIGN alignment program with a PAM120 scoring matrix.

Figures 4A-4D depict an alignment of the nucleotide sequence of the open reading frames of human HtrA (nucleotides 39 to 1478 of SEQ ID NO:10)(SEQ ID NO:) and human HtrA-2 (SEQ ID NO:2). The nucleotide sequences of the open reading frames of human HtrA (SEQ ID NO:) and human HtrA-2 (SEQ ID NO:1) are 62.3% identical. This alignment was performed using the ALIGN alignment program with a PAM120 scoring matrix.

Figures 5A-5B depicts an alignment of the amino acid sequence of human HtrA (SEQ ID NO:11) and the amino acid sequence of human HtrA-2 (SEQ ID NO:3). The amino acid sequences of human HtrA and human HtrA-2 are 56.5% identical. This alignment was performed using the ALIGN alignment program with a PAM120 scoring matrix.

Figures 6A and 6B depict the cDNA sequence of human TANGO 219 (SEQ ID NO:12) and the predicted amino acid sequence of TANGO 219 (SEQ ID NO:14). The open reading frame of SEQ ID NO:12 extends from nucleotide 106 to nucleotide 552 of SEQ ID NO:12 (SEQ ID NO:13).

Figure 7 depicts a hydropathy plot of human TANGO 219. Relatively hydrophobic

regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the amino acid sequence of TANGO 219 is depicted.

Figure 8 depicts the cDNA sequence of mouse HtrA-2 (SEQ ID NO:17) and the predicted amino acid sequence of HtrA-2 (SEQ ID NO:19). The open reading frame of SEQ ID NO: extends from nucleotide 268 to 1311 nucleotide of SEQ ID NO: (SEQ ID NO:19).

Figure 9 depicts the cDNA sequence of mouse TANGO 219 (SEQ ID NO:20) and the predicted amino acid sequence of TANGO 219 (SEQ ID NO:22). The open reading frame of SEQ ID NO: comprises from nucleotide 2 to 370 nucleotide of SEQ ID NO: (SEQ ID NO:21).

Detailed Description of the Invention

The HtrA-2 proteins and nucleic acid molecules comprise a family of molecules having certain conserved structural and functional features. Similarly, the TANGO 219 proteins and nucleic acid molecules comprise a family of molecules having conserved functional and structural features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprises two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. Members of the same family may also have common structural domains.

For example, HtrA-2 proteins and TANGO 219 proteins of the invention have signal sequences. As used herein, a "signal sequence" includes a peptide of at least about 15 or 20 amino acid residues in length which occurs at the N-terminus of secretory and membrane-bound proteins and which contains at least about 70% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 15 to 25 amino acid residues, preferably about 18-22 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. A signal sequence serves to direct

a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, an HtrA-2 protein contains a signal sequence of about amino acids 1 to 17 of SEQ ID NO:3 (SEQ ID NO:5). In another embodiment, a TANGO 219 protein contains a signal sequence of about amino acids 1 to 18 of SEQ ID NO:14 (SEQ ID NO:16). The signal sequence is normally cleaved during processing of the mature protein.

HtrA-2 family members can also include an IGF-binding domain. As used herein, the term "IGF-binding domain" refers to a cysteine rich protein domain that includes about 40-80 amino acid residues, preferably about 50-70 amino acid residues, more preferably about 55-65 amino acid residues, and most preferably about 61 amino acid residues. Typically, an IGF-binding domain is found at the N-terminal half of HtrA-2 and includes a cluster of about 6-15 cysteine residues conserved in IGF binding protein family members, more preferably about 8-10 cysteine residues, and still more preferably about 11 cysteine residues. In addition, an IGF-binding domain includes at least the following consensus sequence: C-Xaa-C-C-Xaa(n1)-C-Xaa-Xaa(n2)-C, wherein C is a cysteine residue, Xaa is any amino acid, n1 is about 1-5 amino acid residues, more preferably about 1-3 amino acid residues, and more preferably 2 amino acid residues in length, and n2 is about 2-10 amino acid residues, more preferably 5-10 amino acid residues, and more preferably 6 amino acid residues in length. In a preferred embodiment, an IGF-binding domain includes at least the following consensus sequence: C-Xaa-C-C-Xaa(n1)-C-A-Xaa(n2)-C, wherein C is a cysteine residue, Xaa is any amino acid, n1 is about 1-5 amino acid residues, more preferably about 1-3 amino acid residues, and more preferably 2 amino acid residues in length, and n2 is about 2-10 amino acid residues, more preferably 5-10 amino acid residues, and more preferably 6 amino acid residues in length.

In one embodiment, an HtrA-2 family member includes an IGF-binding domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 18 to 78 of SEQ ID NO:3, which is the IGF-binding domain of HtrA-2 (this IGF-binding domain is also represented as SEQ ID NO:6). In another embodiment, an HtrA-2 family member includes an IGF-binding domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 18 to 78 of SEQ ID NO:3 (SEQ ID NO:6), includes a conserved cluster of 11 cysteine residues, and an IGF-binding domain consensus sequence as described herein. In yet another embodiment, an HtrA-2 family member includes an IGF-binding domain having an amino acid sequence that is at least

55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 18 to 78 of SEQ ID NO:3 (SEQ ID NO:6), includes a conserved cluster of 11 cysteine residues, an IGF-binding domain consensus sequence as described herein, and has at least one HtrA-2 biological activity as described herein.

In a preferred embodiment, an HtrA-2 family member has the amino acid sequence of SEQ ID NO:3 wherein the cluster of conserved cysteine residues is located within amino acid residues 25 to 76 (at positions, 25, 29, 34, 39, 48, 50, 51, 54, 62, 70, and 76), and the IGF-binding domain consensus sequence is located at amino acid residues 48 to 62.

An HtrA-2 family member can also include a Kazal protease inhibitor domain. As used herein, the term "Kazal protease inhibitor domain" refers to a protein domain that includes about 30-70 amino acid residues, preferably about 40-60 amino acid residues, more preferably about 45-55 amino acid residues, and most preferably about 48 amino acid residues. Typically, a Kazal protease inhibitor domain includes a conserved tyrosine residue and a conserved cluster of about 3-7 cysteine residues, preferably about 4-6 cysteine residues, and still more preferably about 5 cysteine residues. In addition, a Kazal serine protease inhibitor domain includes at least the following consensus sequence: C-Xaa(n1)-C-Xaa(n2)-Y-Xaa(3)-C, wherein C is a cysteine residue, Xaa is any amino acid, n1 is about 4-10 amino acid residues in length, more preferably about 5-8 amino acid residues, and most preferably about 6 amino acid residues in length, n2 is about 4-10 amino acid residues, more preferably about 5-8 amino acid residues, and most preferably about 6 amino acid residues in length, Y is a tyrosine residue, and 3 represents a length of 3 amino acid residues of the type preceding it (in this case 3 of any amino acid (Xaa)).

In one embodiment, an HtrA-2 family member includes a Kazal protease inhibitor domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 79 to 126 of SEQ ID NO:3 (SEQ ID NO:7). In another embodiment, an HtrA-2 family member includes a Kazal protease inhibitor domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 79 to 126 of SEQ ID NO:3 (SEQ ID NO:7), includes a cluster of 5 cysteine residues, and a Kazal protease inhibitor domain consensus sequence as described herein. In yet another embodiment, an HtrA-2 family member includes a Kazal protease inhibitor domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least

about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 79 to 126 of SEQ ID NO:3 (SEQ ID NO:7), includes a cluster of 5 cysteine residues, and a Kazal protease inhibitor domain consensus sequence as described herein, and has at least one HtrA-2 biological activity as described herein.

In a preferred embodiment, an HtrA-2 family member has the amino acid sequence of SEQ ID NO:3 wherein the cluster of 5 cysteine residues is located within amino acid residues 81 to 126 (at positions 81, 83, 90, 101, and 126) and the Kazal protease inhibitor domain consensus sequence is located from amino acid residues 83 to amino acid residue 101.

An HtrA-2 family member can also include a serine protease domain. As used herein, the term "serine protease domain" refers to a protein domain that includes about 180-240 amino acid residues, preferably about 190-230 amino acid residues, more preferably about 205-215 amino acid residues, and most preferably about 208 amino acid residues. In addition, a serine protease domain includes a conserved serine residue, a conserved histidine residue, and a conserved aspartic acid residue in its active site. The conserved histidine, aspartic acid, and serine residues typically appear in the active site within three motifs: 1) a conserved histidine active site motif as follows: Thr-Asn-Xaa-His-Val, where Xaa represents Ala or Asn; 2) a conserved aspartic acid active site motif as follows: Asp-Ile-Ala-Xaa-Ile, where Xaa represents Leu or Thr; and 3) a conserved serine active site motif as follows: Gly-Asn-Ser-Gly-Gly-Xaa-Leu, where Xaa represents Pro or Ala. The conserved histidine active site motif is typically N-terminal to the conserved aspartic acid active site motif, which is N-terminal to the conserved serine active site motif. The histidine and aspartic acid motifs are typically separated from (noninclusive of the last amino acid residue of the first motif and the first residue of the subsequent motif) one another by at least about 15 to 55 amino acid residues, more preferably about 25 to 45 amino acid residues, still more preferably about 30 to 40 amino acid residues, and most preferably about 34 amino acid residues. The aspartic acid and serine motifs are typically separated from (noninclusive of the last amino acid residue of the first motif and the first residue of the subsequent motif) one another by at least about 50 to 90 amino acid residues, more preferably 60 to 80 amino acid residues, still more preferably about 65 to 78 amino acid residues, and most preferably about 71 amino acid residues.

In one embodiment, an HtrA-2 family member includes a serine protease domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 140 to 347 of SEQ ID NO:1, which is

the serine protease domain of HtrA-2 (SEQ ID NO:8). In another embodiment, an HtrA-2 family member includes a serine protease domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 140 to 347 of SEQ ID NO:3 (SEQ ID NO:8) and includes a conserved histidine active site motif, a conserved aspartic acid active site motif, and a conserved serine active site motif as described herein. In yet another embodiment, an HtrA-2 family member includes a serine protease domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 140 to 347 of SEQ ID NO:3 (SEQ ID NO:8), includes a conserved histidine active site motif, a conserved aspartic acid active site motif, and a conserved serine active site motif as described herein, and has at least one HtrA-2 biological activity as described herein.

In a preferred embodiment, an HtrA-2 family member has the amino acid sequence of SEQ ID NO:3 wherein the conserved histidine active site motif is located at amino acid residues 188 to 192 (the histidine residue is at position 191), the conserved aspartic acid active site motif is located at amino acid residues 227 to 231 (the aspartic acid residue is at position 227), and the conserved serine active site motif is located at amino acid residues 303 to 309 (the serine residue is at position 305).

In one embodiment, an HtrA-2 family member can also include a PDZ domain. As used herein, the term "PDZ domain" refers to a protein domain that includes about 70-110 amino acid residues, preferably about 80-100 amino acid residues, more preferably about 87-97 amino acid residues, and most preferably about 92 amino acid residues. Typically, a PDZ domain is located at the C-terminal half of the HtrA-2 protein and includes at least about 3-7 conserved glycine residues, more preferably about 4-6 conserved glycine residues, and most preferably about 5 conserved glycine residues. Typically, a PDZ domain also includes at least the following consensus sequence of G-G-Xaa(n)-D-Xaa(n)-N-G, wherein G is glycine, Xaa is any amino acid, n is about 4-10 amino acid residues in length, more preferably about 5-8 amino acid residues in length, and more preferably about 5-6 amino acid residues in length, and D is aspartic acid.

In one embodiment, an HtrA-2 family member includes a PDZ domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 348 to 439 of SEQ ID NO:3 (SEQ ID NO:9). In another embodiment, an HtrA-2 family member includes PDZ domain having an amino acid

sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 348 to 439 of SEQ ID NO:3 (SEQ ID NO:9) and is located at the C-terminal half of the protein and has a PDZ domain consensus sequence as described herein. In another embodiment, an HtrA-2 family member includes a PDZ domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 348 to 439 of SEQ ID NO:3 (SEQ ID NO:9), is located at the C-terminal half of the protein, includes about 5 conserved glycine residues, and has a PDZ domain consensus sequence as described herein. In yet another embodiment, an HtrA-2 family member includes a PDZ domain having an amino acid sequence that is at least about 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to amino acids 348 to 439 of SEQ ID NO:3 (SEQ ID NO:9), is located at the C-terminal half of the protein, includes about 5 conserved glycine residues, has a PDZ domain consensus sequence as described herein, and has at least one HtrA-2 biological activity as described herein.

In a preferred embodiment, an HtrA-2 family member has the amino acid sequence of SEQ ID NO:3 wherein the PDZ domain is located at the C-terminal half of the protein, from amino acid residues 348 to 439, the PDZ domain consensus sequence is located from amino acid residues 400 to amino acid residue 413, and the conserved glycine residues are located within amino acid residues 358-413 (at positions 358, 385, 400, 401, and 413).

In another embodiment, the signal sequence and the IGF-binding domain of the HtrA-2 family member are adjacent (i.e., there are no intervening residues between the last residue of the signal sequence and the first residue of the IGF-binding domain) to one another. In another example, the IGF-binding domain and Kazal protease inhibitor domain are adjacent (i.e., there are no intervening residues between the last residue of the IGF-binding domain and the first residue of the Kazal protease inhibitor domain) to one another, and the IGF-binding domain is N-terminal to the Kazal protease inhibitor domain. In still another example, the signal sequence is adjacent and N-terminal to the IGF-binding domain, and the Kazal protease inhibitor domain is adjacent to and C-terminal to the IGF-binding domain.

Various features of HtrA-2 and TANGO 219 are summarized below.

Human HtrA-2

A cDNA encoding human HtrA-2 was identified by analyzing the sequences of clones present in an LPS-stimulated osteoblast cDNA library and a prostate stroma cDNA library. This analysis led to the identification of a clone, jthqc058b12, encoding full-length human HtrA-2. The human HtrA-2 cDNA of this clone is 2577 nucleotides long (Figures 1A-1C; SEQ ID NO:1). The open reading frame of this cDNA, nucleotides 222 to 1580 of SEQ ID NO:1 (SEQ ID NO:2), encodes a 453 amino acid secreted protein (Figures 1A-1C; SEQ ID NO:3).

In one embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 278 is an guanine (G) (SEQ ID NO:1). In this embodiment, the amino acid at position 19 is glutamate (E)(SEQ ID NO:3) In another embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 278 is a cytosine (C)(SEQ ID NO:24). In this embodiment, the amino acid at position 19 is aspartate (D)(SEQ ID NO:25) In another embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 395 is guanine (G)(SEQ ID NO:1). In this embodiment, the amino acid at position 58 is glutamate (E)(SEQ ID NO:3). In another embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 395 is cytosine (C)(SEQ ID NO:26). In this embodiment, the amino acid at position 58 is aspartate (D)(SEQ ID NO:27). In another embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 401 is guanine (G)(SEQ ID NO:1). In this embodiment, the amino acid at position 60 is glutamate (E)(SEQ ID NO:3). In another embodiment of a nucleotide sequence of human HtrA-2, the nucleotide at position 401 is cytosine (C)(SEQ ID NO:28). In this embodiment, the amino acid at position 60 is aspartate (D)(SEQ ID NO:29).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human HtrA-2 includes a 17 amino acid signal peptide (amino acid 1 to about amino acid 17 of SEQ ID NO:3)(SEQ ID NO:5) preceding the mature HtrA-2 protein (corresponding to about amino acid 18 to amino acid 453 of SEQ ID NO:3)(SEQ ID NO:4). The HtrA-2 protein molecular weight is 48.6 kDa prior to the cleavage of the signal peptide, 47.0 kDa after cleavage of the signal peptide.

HtrA-2 includes a IGF binding domain (about amino acids 18 to 78 of SEQ ID NO:3; SEQ ID NO:6), a Kazal protease inhibitor domain (about amino acids 79 to 126 of SEQ ID NO:3; SEQ ID NO:7), a serine protease domain (about amino acids 140 to 347 of SEQ ID NO:3; SEQ ID NO:8), and a PDZ domain (about amino acids 348-439 of SEQ ID NO:3; SEQ ID NO:9).

Figures 3A-3G shows an alignment of the human HtrA-2 full length nucleic acid

sequence (SEQ ID NO:1) with the human HtrA full length nucleic acid sequence (SEQ ID NO:10). Figures 4A-4D shows an alignment of the human HtrA-2 nucleotide coding region (SEQ ID NO:2) with the human HtrA nucleotide coding region. Figures 5A-5B shows an alignment of the human HtrA-2 protein sequence (SEQ ID NO:3) with the human HtrA protein sequence (SEQ ID NO:12). As shown in Figures 5A-5B, the human HtrA-2 signal sequence is represented by amino acids 1-17 (and encoded by nucleotides 222-272 of SEQ ID NO:1), and the human HtrA signal sequence is represented by amino acids 1-22 (and encoded by nucleotides 39-103 of SEQ ID NO:12). The human HtrA-2 IGF-binding domain sequence (SEQ ID NO:6) is represented by amino acids 18-78 (and encoded by nucleotides 273-455 of SEQ ID NO:1)(SEQ ID NO:), and the human HtrA IGF-binding sequence is represented by amino acids 37-94 (and encoded by nucleotides 147-320 of SEQ ID NO:12)(SEQ ID NO:). The human HtrA-2 Kazal protease inhibitor domain sequence (SEQ ID NO:7) is represented by amino acids 79-126 (and encoded by nucleotides 456-599 of SEQ ID NO:1)(SEQ ID NO:), and the human HtrA Kazal protease inhibitor domain sequence is represented by amino acids 110-155 (and encoded by nucleotides 366-503 of SEQ ID NO:12)(SEQ ID NO:). The human HtrA-2 serine protease domain sequence (SEQ ID NO:8) is represented by amino acids 140-347 (and encoded by nucleotides 639-1262 of SEQ ID NO:1)(SEQ ID NO:), and the human HtrA serine protease domain sequence is represented by amino acids 140-369 (and encoded by nucleotides 456-1145 of SEQ ID NO:12)(SEQ ID NO:). The human HtrA-2 PDZ domain sequence (SEQ ID NO:9) is represented by amino acids 348-439 (and encoded by nucleotides 1263-1538 of SEQ ID NO:1)(SEQ ID NO:), and the human HtrA PDZ domain sequence is represented by amino acids 370-465 (and encoded by nucleotides 1146-1433 of SEQ ID NO:12)(SEQ ID NO:).

Figures 3A-3G and Figures 4A-4D show that there is an overall 50.9% identity between the full length human HtrA-2 nucleic acid molecule and the full length human HtrA nucleic acid molecule, and an overall 62.3% identity between the open reading frame of human HtrA-2 nucleic acid molecule (SEQ ID NO:2) and the open reading frame of the human HtrA nucleic acid molecule (SEQ ID NO:11), respectively. The amino acid alignment in Figures 5A-5B shows a 56.5% overall amino acid sequence identity between human HtrA-2 (SEQ IDNO:3) and human HtrA (SEQ ID NO:12).

Clone EpT214, which encodes human HtrA-2, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on September 25, 1998 and assigned Accession Number 98899. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made

merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 2 depicts a hydropathy plot of human HtrA-2. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace.

Northern analysis of HtrA-2 expression in human tissues showed that an approximately 2.6 kB transcript is expressed in human adult heart, skeletal muscle, lung, pancreas, and placenta. No expression was detected in the kidney or brain. In comparison, Northern analysis of HtrA expression in human tissue (Zumbrunn, et al. (1996) *FEBS Lett.* 398:187-192) indicated that an approximately 2.3 kB transcript is strongly expressed in human placenta, moderately expressed in human brain, liver, and kidney, and weakly expressed in human lung, skeletal muscle, heart, and pancreas.

Library Array Expression: Expression of human HtrA-2 mRNA was detected by a library array procedure. Briefly, this entailed preparing a PCR mixture by adding standard reagents (e.g., Taq Polymerase, dNTPs, and PCR buffer) a vector primer, a primer internal to the gene of interest, and an aliquot of a library in which expression was to be tested. This procedure was performed with many libraries at a time in a 96 well PCR tray, with 80 or more wells containing libraries and a control well in which the above primers were combined with the clone of interest itself. The control well served as an indicator of the fragment size to be expected in the library wells, in the event the clone of interest was expressed within. Amplification was performed in a PCR machine, employing standard PCR conditions for denaturing, annealing, and elongation, and the resultant mixture was mixed with an appropriate loading dye and run on an ethidium bromide-stained agarose gel. The gel was later viewed with UV light after the DNA loaded within its lanes had time to migrate into the gels. Lanes in which a band corresponding with the control band was visible indicated the libraries in which the clone of interest was expressed.

Expression was detected in human umbilical endothelial cells and human lean subcutaneous adipose tissue. No expression was detected in kidney, testes, prostate, HMC-1 control (mast cell line), fetal dorsal spinal cord, human colon to liver metastasis, erythroblasts from CD34+ blood, human spinal cord (ION 3), HUVEC TGF-B (human umbilical endothelia), HUVEC (human umbilical endothelia), Brain, K563 (red blood cell line), uterus, Hep-G2 (human insulinoma), human normal colon, human colon to liver metastasis, skin, HUVEC controls (umbilical endothelial cells), human colon (inflammatory bowel disease), melanoma (G361 cell line), adult bone marrow CD34+ cells, HPK, human

lung, mammary gland, normal breast epithelium, colon to liver metastasis (CHT128), normal breast, bone marrow (CD34+), WI38 (human embryonic lung), Th1 cells, HUVEC untreated (umbilical endothelium), uterus, liver, spleen, normal human Ovarian Epithelia, colon to liver metastasis (CHT133), PTH-treated osteoblasts, ovarian ascites, lung squamous cell carcinoma (MDA 261), Th2 cells, IBD colon (WUM 23), thymus, heart, small intestine, normal megakaryoctyes, colon carcinoma (NDR109), lung adenocarcinoma (PIT245), IBD colon (WUM6), brain-subcortical white matter (ION2), prostate tumor xenograft A12, trigeminal ganglia, 9 week fetus, retinal pigmentosa epithelia, bone marrow, colon arcinoma (NDR103), lung squamous cell carcinoma (PIT299), cervical cancer, normal prostate, prostate tumor, xenograft K10, lumbrosacaral spinal cord, A549 control, stomach, retina, Th-1 induced T cell, colon carcinoma (NDR82), d8 dendritic cells, spinal cord, ovarian epithelial tumor, prostate cancer to liver metastasis JHH3, lumbrosacaral dorsal root ganglia, salivary gland, skeletal muscle, HMC-1 (human mast cell line), Th-2 induced T-cell, colon carcinoma (NDR097), megakaryocytes, dorsal root ganglia (ION 6, 7, 8), HUVEC L-NAME (umbilical endothelia), prostate cancer to liver metastasis JHH4, dorsal root ganglia (ION 6, 7, 8), HMVEC: microvascular endothelial cells, fetal brain, bronchial epithelium mix, mesangial cells, fetal heart, LPS-stimulated 24 hours Osteoblasts, cervical carcinoma A2780 WT cell line, UCLA-lung carcinoma R (carcinoma (Resistant to drug treatment)), erythroleukemia cells, trachea, testes, placenta, HUVEC: umbilical vein endothelial cells, bronchial epithelium, congestive heart failure, bladder carcinoma T24 cell line Ctl., mammary gland, burkitt's lymphoma, cervical carcinoma A2780 ADR cell line (drug resistant), UCLA-lung carcinoma S (carcinoma (sensitive to drug treatment)), embryonic keratinocytes, cervix carcinoma ME180 IL-1, testes, mammary gland, HL60/S, astrocytes, cerebellum, bladder carcinoma T24 Tr., natural killer cells, fetal spleen, Prostate, fetal fibroblast, SCC25 CDDP - tongue squamous carcinoma, cervix carcinoma, ME 180 control, RAJI- human burkitt's lymphoma B cell, small intestine, U937/A10P10, prostate epithelium, pituitary, prostate fibroblast, congestive heart failure, uterine smooth Muscle, treated, esophagus, p65 IL-1, SCC25 WT- tongue squamous cell carcinoma, MCP-1 mast cell line, ST486 (Lymphoma B cell), fetal liver, U937/ A10p50, primary osteoblast, Aortic endothelial cells, bone marrow, prostate smooth muscle, umbilical smooth muscle, treated, fetal liver, lung carcinoma A549 control, fetal hypothalamus, HPK II keratinocyte cell line, HL60 (acute Promyelocytic Leukemia), skeletal muscle, CaCo, keratinocytes, fetal Kidney, congestive heart failure, thyroid, bronchial smooth muscle, fetal skin, A549IL-1, T cells, CD3 treated, lung, umbilical smooth muscle, treated, stomach, HeLa cells, melanocytes, fetal liver, adrenal gland, LPS-stimulated osteoblasts, 1 hour, WT LNCap+ casodex, fetal

adrenal gland, fetal testes, T cells, CD3 IL-4/ IL-10 treated, heart, uterine smooth muscle, spleen, HL60/Adr, coronary smooth muscle cells, fetal lung, fetal thymus, LPS-stimulated 6 hour osteoblasts, WT LNCap+ Testosterone, midterm placenta, pulmonary artery smooth muscle, T cells, CD3 IFNg/TFNa treated, fetal brain, and liver.

Mouse HtrA-2

A mouse homolog of human HtrA-2 was identified. A cDNA encoding mouse HtrA-2 was identified by analyzing the sequences of clones present in a mouse cDNA library. This analysis led to the identification of a clone, Atmx2143, encoding full-length mouse HtrA-2. The murine HtrA-2 cDNA of this clone is 1563 nucleotides long (Figure 8; SEQ ID NO:18). The open reading frame of this cDNA, comprises nucleotides 268 to 1311 of SEQ ID NO:18 (SEQ ID NO:19), and encodes a 349 amino acid secreted protein (Figure 8; SEQ ID NO:20).

In one embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 396 is an guanine (G) (SEQ ID NO:18). In this embodiment, the amino acid at position 43 is glutamate (E)(SEQ ID NO:20) In another embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 396 is a cytosine (C)(SEQ ID NO:30). In this embodiment, the amino acid at position 43 is aspartate (D)(SEQ ID NO:31) In another embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 426 is guanine (G)(SEQ ID NO:18). In this embodiment, the amino acid at position 53 is glutamate (E)(SEQ ID NO:20). In another embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 426 is cytosine (C)(SEQ ID NO:32). In this embodiment, the amino acid at position 53 is aspartate (D)(SEQ ID NO:33). In another embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 498 is guanine (G)(SEQ ID NO:18). In this embodiment, the amino acid at position 77 is glutamate (E)(SEQ ID NO:20). In another embodiment of a nucleotide sequence of mouse HtrA-2, the nucleotide at position 498 is cytosine (C)(SEQ ID NO:34). In this embodiment, the amino acid at position 77 is aspartate (D)(SEQ ID NO:35).

HtrA-2 mRNA expression in mouse: In situ tissue screening was performed on mouse adult and embryonic tissue to analyze for the expression of mouse HtrA-2 mRNA. In summary, adult expression was highest in the bladder and present to a lesser extent in heart, muscle, and colon. Signal in these tissues was ubiquitous. All other adult tissues showed no specific signal above background. Expression at embryonic day 13.5, the earliest age tested, was observed in the stomach and brain. Expression pattern in brain was punctate, broadly distributed, and sparse. Beginning at E14.5 ubiquitous expression was also

observed in skeletal muscle, diaphragm, intestine, and lung. This pattern continues until postnatal day 1.5, when expression was also apparent in the renal medulla. There was high background in what appears to be cartilage. The antisense probe showed a stronger signal in this tissue with a more extensive pattern. In particular, with respect to adult mouse expression, expression was ubiquitous in each of the bladder, heart, skeletal muscle, and colon. No expression was detected in the following tissues: lung, brain, placenta, liver, pancreas, thymus, eye, kidney, and the small intestine. With respect to expression in the embryonic mouse, the following results were obtained: At E13.5, expression was detected in the stomach and brain. Signal was also observed in the limbs and vertebrae with the sense probe. This signal was much higher and more extensive with the antisense probe. At E14.5, E15.5, E16.5, E18.5, and P1.5, a signal was punctuate in the brain, strong in renal medulla and absent from liver. Most other tissues had low level ubiquitous expression to some degree.

Uses of HtrA-2 Nucleic acids, Polypeptides, and Modulators Thereof

As HtrA-2 was originally found in an LPS-treated osteoblast library and is homologous to HtrA, mRNA levels of which are known to be elevated in cartilage from individuals with osteoarthritis, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat bone and/or cartilage associated diseases or disorders. Examples of bone and/or cartilage diseases and disorders include bone and/or cartilage injury due to for example, trauma (e.g., bone breakage, cartilage tearing), degeneration (e.g., osteoporosis), degeneration of joints, e.g., arthritis, e.g., osteoarthritis, and bone wearing.

In addition, as HtrA-2, like HtrA, is highly expressed in heart tissue, and includes an IGF-binding domain, and thus likely has a role in modulating IGF function (e.g., IGF is involved in cardiac hyperplasia), HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat disorders of the cardiovascular system. Examples of disorders of the cardiovascular system include disorders associated with a lack of nutrients to the heart muscle, e.g., ischemic heart disease, atherosclerosis, disorders associated with abnormal morphology of the heart, e.g., cardiac hyperplasia, and disorders associated with abnormal contraction of the heart, e.g., ventricular fibrillation.

The presence of an IGF binding domain in HtrA-2 also suggests that HtrA-2 can modulate IGF function and thereby be used to treat IGF associated disorders. IGFs are known to be involved in the overall cellular growth of embryos and organs of mammals. When existing at excessive levels, however, IGFs can cause somatic overgrowth which leads to conditions such as visceromegaly, placentomegaly, cardiac and adrenal defects, and

Beckwith-Weidermann syndrome. Thus, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat IGF-associated disorders as described above. In addition, as IGF can cause increased cell proliferation, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat proliferative disorders, e.g., cancer, e.g., cancer of a cell or tissue in which HtrA-2 is expressed.

The presence of a Kazal protease inhibitor domain in HtrA-2 also indicates that HtrA-2 can function in a similar manner as other proteins containing a Kazal protease inhibitor domain. For example, follistatin includes a Kazal protease inhibitor domain. Follistatin regulates the availability of growth factors and embryonic growth, and thus modulators thereof can be used to treat disorders involving abnormal cellular migration, proliferation, and differentiation. Similarly, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat disorders involving abnormal cellular migration, proliferation (e.g., cancer), and/or differentiation, and/or follistatin-associated disorders.

As HtrA-2 includes a serine protease domain, it can act as a serine protease. Thus, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat disorders involving abnormal serine protease function. For example, it is known that serine protease inhibitors are abundant in plaques found in Alzheimer's patients, and may be responsible for preventing some types of metalloproteinase from breaking down the beta-amyloid proteins that make up these plaques. Thus, modulation of the HtrA-2 serine protease activity may modulate formation of Alzheimer's plaques. Consequently, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat Alzheimer's disease.

The presence of a PDZ domain in HtrA-2 suggests that HtrA-2 functions in a manner similar to other PDZ-containing proteins. For example, PDZ domains typically bind other proteins at their carboxyl termini in a sequence-specific manner.

Human HtrA-2 nucleic acids, proteins, and modulators thereof can also be used to treat neurological disorders. Examples of such neurological disorders include disorders due to nerve damage (e.g., nerve damage due to stroke) and neurodegenerative diseases (e.g., Alzheimer's disease, multiple sclerosis, Huntington's disease, and Parkinson's disease). In addition, human HtrA-2 nucleic acids, polypeptides, and modulators thereof can be used to treat neurodegeneration associated with Alzheimer's disease, frontal lobe dementia, cortical lewy body disease, dementia of Parkinson's disease, acute and chronic phases of degeneration following stroke or head injury, neuronal degeneration found in motor neuron disease, AIDS dementia and chronic epilepsy.

HtrA-2 , like HtrA, can likely interact with a normal or mutated gene product of a human presenilin gene (e.g., human presenilin-1 (PS-1), e.g., the hydrophobic loop domain

between transmembrane domains 1 and 2 of PS-1). As mutations in the human PS-1 gene lead to Familial Alzheimer's disease (see PCT Publication Number WO98/01549, the contents of which are incorporated by reference), and HtrA-2 can interact with PS-1 and thus modulate PS-1 function, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat Alzheimer's disease and physiological functions associated with Alzheimer's disease.

The PS-1 gene product may also be a receptor or channel protein, mutations in which have been causally related to neurological disorders whose pathology does not represent Alzheimer's disease. Thus, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat non-Alzheimer's neurological disorders as well (e.g., malignant hyperthermia, hyperkalemic periodic paralysis).

HtrA-2 nucleic acids, proteins, and modulators thereof can also be used to treat disorders of the cells and tissues in which it is expressed. As HtrA-2 is expressed in colon, bladder, skeletal muscle, lung, pancreas, and placenta, HtrA-2 nucleic acids, proteins, and modulators thereof can be used to treat disorders of these cells, tissues, or organs, e.g., colon cancer and colonic volvulus, diverticula, cystitis, urinary tract infection, bladder cancer, muscular dystrophy, stroke, muscular atrophy, trichinosis, lung cancer, cystic fibrosis, rheumatoid lung disease, pancreatic cancer, diabetes, pancreatitis, and various placental disorders.

Because HtrA-2 was expressed in adipose tissue, HtrA-2 nucleic acids, proteins and modulators thereof can be utilized to modulate adipocyte function and adipocyte-related processes and disorders such as, e.g., obesity, regulation of body temperature, lipid metabolism, carbohydrate metabolism, body weight regulation, obesity, anorexia nervosa, diabetes mellitus, unusual susceptibility or insensitivity to heat or cold, arteriosclerosis, atherosclerosis, and disorders involving abnormal vascularization, e.g., vascularization of solid tumors. Additionally, such molecules can be used to treat disorders associated with abnormal fat metabolism, e.g., cachexia. In another example, such molecules can be used to treat disorders associated with abnormal proliferation of these tissues, e.g., cancer, e.g., breast cancer or liver cancer.

Human TANGO 219

A cDNA encoding human TANGO 219 was identified by analyzing the sequences of clones present in a prostate stroma cDNA library. This analysis led to the identification of a clone, jthqc101c05, encoding full-length human TANGO 219. The human TANGO 219 cDNA of this clone is 1268 nucleotides long (Figure 6; SEQ ID NO:13). The open reading

frame of this cDNA, nucleotides 106 to 552 of SEQ ID NO:12 (SEQ ID NO:14), encodes a 149 amino acid secreted protein (Figure 6; SEQ ID NO:15).

In one embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 186 is an guanine (G) (SEQ ID NO:13). In this embodiment, the amino acid at position 27 is glutamate (E)(SEQ ID NO:15) In another embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 186 is a cytosine (C)(SEQ ID NO:36). In this embodiment, the amino acid at position 27 is aspartate (D)(SEQ ID NO:37) In another embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 261 is guanine (G)(SEQ ID NO:13). In this embodiment, the amino acid at position 52 is glutamate (E)(SEQ ID NO:15). In another embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 261 is cytosine (C)(SEQ ID NO:38). In this embodiment, the amino acid at position 52 is aspartate (D)(SEQ ID NO:39). In another embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 309 is adenine (A)(SEQ ID NO:13). In this embodiment, the amino acid at position 68 is glutamate (E)(SEQ ID NO:15). In another embodiment of a nucleotide sequence of human TANGO 219, the nucleotide at position 309 is cytosine (C)(SEQ ID NO:40). In this embodiment, the amino acid at position 68 is aspartate (D)(SEQ ID NO:41).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 219 includes an 18 amino acid signal peptide (amino acid 1 to about amino acid 18 of SEQ ID NO:16)(SEQ ID NO:17) preceding the mature TANGO 219 protein (corresponding to about amino acid 19 to amino acid 149 of SEQ ID NO:14)(SEQ ID NO:15). The TANGO 219 protein molecular weight is 17.0 kDa prior to the cleavage of the signal peptide, 14.8 kDa after cleavage of the signal peptide.

Northern analysis of TANGO 219 in human tissues revealed that an approximately 1.25 kB transcript is weakly expressed in the heart and not in other tissues tested such as, for example, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, stomach, thyroid, spinal cord, lymph node, adrenal gland, trachea, bone marrow, spleen, thymus, prostate, testes, ovary, small intestine, colon, and peripheral blood leukocytes.

The human gene for TANGO 219 was mapped on radiation hybrid panels to the long arm of chromosome 5, in the region q21-22. Flanking markers for this region are NIB916 and D5S492. The MANA2 (mannosidase, type2), APC (adenomatous polyposis coli), PST (polysialytransferase), CAST (calpastatin) genes also map to this region of the human chromosome. The LGMD1A (limb girdle muscular dystrophy) loci also maps to this region of the human chromosome. This region is syntenic to mouse chromosomes 11 and 18

. The Q (quinky), pdw (proportional dwarf), and lyl1 (lymphoblastomic leukemia) loci also map to this region of the mouse chromosome. The Chr.11-fer (protein kinase, testis specific) Chr.18-mcc (mutated in colorectal cancers), pk (plucked), donl (divergent of neuregulin 1) genes also map to this region of the mouse chromosome.

Clone EpT219, which encodes human TANGO 219, was deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on September 25, 1998 and assigned Accession Number 98899. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 7 depicts a hydropathy plot of human TANGO 219. Relatively hydrophobic regions of the protein are shown above the horizontal line, and relatively hydrophilic regions of the protein are below the horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Mouse TANGO 219

A mouse homolog of human TANGO 219 was identified. A cDNA encoding mouse TANGO 219 was identified by analyzing the sequences of clones present in a mouse cDNA library. This analysis led to the identification of a clone encoding mouse TANGO 219. The murine TANGO 219 cDNA of this clone is 397 nucleotides long (Figure 9; SEQ ID NO:21). The open reading frame of this cDNA, comprises nucleotides 2 to 370 of SEQ ID NO:(SEQ ID NO:22), and encodes a secreted protein comprising the 123 amino acid sequence depicted in Figure 9 (SEQ ID NO:23).

In one embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 127 is an guanine (G) (SEQ ID NO:21). In this embodiment, the amino acid at position 42 is glutamate (E)(SEQ ID NO:23) In another embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 127 is a cytosine (C)(SEQ ID NO:42). In this embodiment, the amino acid at position 42 is aspartate (D)(SEQ ID NO:43) In another embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 154 is guanine (G)(SEQ ID NO:21). In this embodiment, the amino acid at position 51 is glutamate (E)(SEQ ID NO:23). In another embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 154 is cytosine (C)(SEQ ID NO:44). In this embodiment, the amino acid at position 51 is aspartate (D)(SEQ ID NO:45).

In another embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 226 is adenine (A)(SEQ ID NO:21). In this embodiment, the amino acid at position 75 is glutamate (E)(SEQ ID NO:23). In another embodiment of a nucleotide sequence of mouse TANGO 219, the nucleotide at position 226 is cytosine (C)(SEQ ID NO:46). In this embodiment, the amino acid at position 75 is aspartate (D)(SEQ ID NO:47).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that mouse TANGO 219 includes a 40 amino acid signal peptide (amino acid 1 to about amino acid 40 of SEQ ID NO:23)(SEQ ID NO:34) preceding the mature mouse TANGO 219 protein (corresponding to about amino acid 41 to amino acid 123 of SEQ ID NO:23)(SEQ ID NO:35). The TANGO 219 protein molecular weight is 17.0 kDa prior to the cleavage of the signal peptide, 14.8 kDa after cleavage of the signal peptide.

Uses of TANGO 219 Nucleic acids, Polypeptides, and Modulators Thereof

As TANGO 219 was originally found in a prostate stroma library, TANGO 219 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function, e.g., secretory activity, of the prostate cells. Such molecules can also be useful for treatment of prostate diseases or disorders, e.g., acute or chronic prostatitis (and the resulting urinary tract infection), benign nodular enlargement, and prostatic carcinoma.

As TANGO 219 appears to be preferentially expressed in the heart and not in many other tissues, TANGO 219 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function of cardiac cells. Thus TANGO 219 plays a role in regulating cardiac function, including modulating cardiac rhythm and strength of contraction, and the heart's response to stress. Such molecules can also be useful for treatment of heart diseases or conditions, e.g., ischemic heart disease or atherosclerosis, or cerebrovascular accidents, or more particularly, treating or preventing conditions involving heart contraction, and the impulse generating nodes and cardiac muscle cells, e.g., ventricular fibrillation or myocardial infarction.

Tables 1, and 2 below provide a summary of the sequence information for human HtrA-2 and human TANGO 219.

Table 3 below provides a summary of the sequence information for mouse HtrA-2 and mouse TANGO 219.

TABLE 1: Summary of Human HtrA-2 and Human TANGO 219 Sequence Information

Gene	cDNA	ORF	Protein	Figure	ATCC Accession No.
HUMAN HtrA-2	SEQ ID NO:1	SEQ ID NO:2	SEQ ID NO:3	Figures 1A-1C	98899
HUMAN TANGO 219	SEQ ID NO:13	SEQ ID NO:14	SEQ ID NO:15	Figures 6A-6B	98899

TABLE 2: Summary of Domains of Human HtrA-2 and Human TANGO 219 Proteins

Protein	Mature Protein	Signal Sequence	IGF Binding	Kazal	Serine Protease	PDZ
HUMAN HtrA-2	aa 18-453 of SEQ ID NO:3 (SEQ ID NO:4)	aa 1-17 of SEQ ID NO:3 (SEQ ID NO:5)	aa 18-78 of SEQ ID NO:3 (SEQ ID NO:6)	aa 79-126 of SEQ ID NO:3 (SEQ ID NO:7)	aa 140-347 of SEQ ID NO:3 (SEQ ID NO:8)	aa 348-439 of SEQ ID NO:3 (SEQ ID NO:9)
HUMAN TANGO 219	aa 19-149 of SEQ ID NO:14 (SEQ ID NO:16)	aa 1-18 of SEQ ID NO:14 (SEQ ID NO:17)				

TABLE 3: Summary of Mouse HtrA-2 and Human TANGO 219 Sequence Information

Gene	cDNA	ORF	Protein	Figure	ATCC Accession No.
MOUSE HtrA-2	SEQ ID NO:18	SEQ ID NO: 19	SEQ ID NO:20	Figure 8	
MOUSE TANGO 219	SEQ ID NO:21	SEQ ID NO:22	SEQ ID NO:23	Figure 9	

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid

molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the

invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologues in other cell types, e.g., from other tissues, as well as homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 or of a naturally occurring mutant of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, e.g., detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion" of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NO:2,14, 19, or 22, expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO:2, 14, 19, or 22.

In addition to the nucleotide sequences of SEQ ID NO:2, 14, 19, and 22, it will be

appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (e.g., the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. For example, TANGO 219 has been mapped to chromosome 5, between flanking markers NIB916 and D5S492, and therefore, TANGO 219 family members can include nucleotide sequence polymorphisms (*e.g.*, nucleotide sequences that vary from SEQ ID NO:) that map to this chromosome 5 region (*i.e.*, between markers NIB916 and D5S492). As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologues), which have a nucleotide sequence which differs from that of the human protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1290) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding

sequence, of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, or a complement thereof.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45EC, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65EC. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). Representative species that hybridize under such conditions to one or more of the sequences above include, but are not limited to, SEQ ID NO:24, 26, and 28, which in particular hybridize to the human HtrA-2 sequences listed above (SEQ ID NOs:1 and 2). Representative species that hybridize under such conditions to one or more of the sequences above include, but are not limited to, SEQ ID NO:30, 32, and 34, which in particular hybridize to the mouse HtrA-2 sequences listed above (SEQ ID NOs:18 and 20). Representative species that hybridize under such conditions to one or more of the sequences above include, but are not limited to, SEQ ID NO:36, 38, and 40, which in particular hybridize to the human TANGO 219 sequences listed above (SEQ ID NOs:13 and 14). Representative species that hybridize under such conditions to one or more of the sequences above include, but are not limited to, SEQ ID NO:42, 44, and 46, which in particular hybridize to the mouse TANGO 219 sequences listed above (SEQ ID NOs: 21 and 22).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved

among homologues of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologues of various species (e.g., murine and human) may be essential for activity and thus would not be likely targets for alteration. For example, representative species of the human HtrA-2 presented for illustrative purposes only and not by way of limitation, include but are not limited to, SEQ ID NOs:25, 27, and 29. For example, representative species of the mouse HtrA-2 presented for illustrative purposes only and not by way of limitation, include but are not limited to, SEQ ID NOs:31, 33, and 35. For example, representative species of the human TANGO 219 presented for illustrative purposes only and not by way of limitation, include but are not limited to, SEQ ID NOs:37, 39, and 41. For example, representative species of the mouse TANGO 219 presented for illustrative purposes only and not by way of limitation, include but are not limited to, SEQ ID NOs:43, 45, and 47.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NO:3 or 14, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:3, 15, 20, or 23.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Preferably such variant proteins retain or exhibit at least one structural or biological activity of the polypeptides of the invention. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine,

methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein:protein interactions with proteins in a signaling pathway of the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention; or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet another preferred embodiment, the mutant polypeptide can be assayed for the ability to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides or more in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil,

5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run

parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al. (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be

performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra,* and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication

No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.,* Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (*see, e.g.,* Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (e.g., the amino acid sequence shown in any of SEQ ID NO:3, 4,

15, 16, 20, 23, 24, or 35), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID NO:3, 4, 15, 16, 20, 23, 24, or 35. Other useful proteins are substantially identical (e.g., at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID NO:3, 4, 15, 16, 20, 23, 24, or 35 and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) x 100). In one embodiment, the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino

acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example,

useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g. promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence of a polypeptide of the invention (SEQ ID NO:5, 17 and 34) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal

sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, e.g., promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal sequence on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A

variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.,* Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA 89*:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or,

alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NO:3, 15, 20, or 23, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Figures 2 and 7 are hydropathy plots of the proteins of the invention. These plots or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, recombinantly expressed or chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen

compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies specific for a protein or polypeptide of the invention can be selected for (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*,

using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816397, which are incorporated herein by reference in their entirety.) Humanized antibodies are antibody molecules from non-human species having one or more complementarily determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.*, Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et

al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a mouse antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al. (1994) *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (*e.g.*, monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable

substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"),

granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, 23, an amino acid sequence encoded by the HtrA-2 cDNA of a clone deposited as ATCC Accession No. 98899, or an amino acid sequence encoded by the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs: 3, 15, 20, 2321; an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, 23, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the HtrA-2 cDNA of a clone deposited as ATCC Accession No. 98899, or the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, or 23, an amino acid sequence encoded by the HtrA-2 cDNA of a clone deposited as ATCC Accession No. 98899, or an amino acid sequence encoded by the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, or 23; an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID Nos:3, 15, 20, or 23, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos:1, 2, 13, 14, 18, 19, 21, 22, the HtrA-2 cDNA of a clone deposited as ATCC Accession No. 98899, or the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID Nos:3, 15, 20, or 23, an amino acid sequence encoded by the HtrA-2 cDNA of a clone deposited as ATCC Accession No. 98899, or an amino acid sequence encoded by the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, or 23; an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:3, 15, 20, or 23, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos:1, 2, 13, 14, 18, 19, 21, 22, the HtrA-2 cDNA of a clone deposited as ATCC Accession No.

98899, or the TANGO 219 cDNA of a clone deposited as ATCC Accession No. 98899, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes HtrA-2 and TANGO 219, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immungen comprises an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NO:3, 15, 20, or 23, an amino acid sequence encoded by the HtrA-2 cDNA of a clone deposited as ATCC accession No. 98899, or an amino acid sequence encoded by the TANGO 219 cDNA of a clone deposited as ATCC accession No. 98899 ; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NO:3, 15, 20 or 23, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NO:3, 15, 20 or 23, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the HtrA-2 cDNA of a clone deposited as ATCC accession No. 98899, or the TANGO 219 cDNA of a clone deposited as ATCC accession No. 98899 , or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. After immunization, a sample is collected from the mammal that contains an antibody that specifically recognizes GPVI. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be

further purified from the sample using techniques well known to those of skill in the art. The method can further comprise producing a monoclonal antibody-producing cell from the cells of the mammal. Optionally, antibodies are collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press,

San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (e.g., *E. coli*) or eukaryotic cells (e.g., insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a

coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra.*

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv.*

Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the beta-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (e.g., *E. coli*) or eukaryotic cell (e.g., insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via

conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous (*e.g.*, HtrA-2 and TANGO 219) nucleic acid within a cell, cell line or microorganism may be modified by inserting a DNA regulatory element heterologous to the endogenous gene of interest into the genome of a cell, stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous gene (*e.g.*, HtrA-2 and TANGO 219) and controls, modulates or activates the endogenous gene. For example, endogenous HtrA-2 and TANGO 219 which are normally "transcriptionally silent", *i.e.*, HtrA-2 and TANGO 219 genes which are normally not expressed, or are expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, transcriptionally silent, endogenous HtrA-2 and TANGO 219 genes may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with and activates expression of endogenous HtrA-2 and TANGO 219 genes, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in

culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequence encoding a polypeptide of the invention has been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to

particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see, e.g.*, Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (*see, e.g.*, Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (*see, e.g.*, Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in

their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication NOS. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.*, Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for

modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL? (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and

antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be

permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to

enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (*see, e.g.,* Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g.,retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). For example, the HtrA-2 polypeptides of the invention can to used to, for example (i) mediate protein-protein interactions; (ii) modulate cellular migration, proliferation, and differentiation; (iii) modulate proteolytic enzyme activity; and/or (iv) modulate protein degradation. The isolated nucleic acid molecules of the invention can be used to express proteins (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (e.g., in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the

invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOS. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (e.g., a polypeptide of the invention) binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected

protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a

target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma

Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or

protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) *Science* 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a

mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

In the instant case, the human gene for TANGO 219 was mapped on radiation hybrid panels to the long arm of chromosome 5, in the region q21-22. Flanking markers for this region are NIB916 and D5S492. The MANA2 (mannosidase, type2), APC (adenomatous polyposis coli), PST (polysialytransferase), CAST (calpastatin) genes also map to this region of the human chromosome. The LGMD1A (limb girdle muscular dystrophy) loci also maps to this region of the human chromosome. This region is syntenic to mouse chromosomes 11 and 18 . The Q (quinky), pdw (proportional dwarf), and lyl1 (lymphoblastomic leukemia) loci also map to this region of the mouse chromosome. The Chr.11-fer (protein kinase, testis specific) Chr.18-mcc (mutated in colorectal cancers), pk (plucked), donl (divergent of neuregulin 1) genes also map to this region of the mouse chromosome.

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 or 12 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:2 or 13 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing

another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, e.g., fragments derived from noncoding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

C. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of a polypeptide of the invention, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide of the invention. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, mutations in a gene of the invention can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with aberrant expression or activity of a polypeptide of the invention.

Another aspect of the invention provides methods for expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide of the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that

individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent).

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of a polypeptide of the invention in clinical trials. These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NO: 1, 2, 13, 14, 18, 19, 21, 22, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells

and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (e.g., a proliferative disorder, e.g., psoriasis or cancer). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at

risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention, e.g., a neurological disorder, e.g., Alzheimer's disease, a bone disorder, or a cardiovascular disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample

obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (e.g., wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of a the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and

4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes

can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a selected sequence, e.g., a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a `GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective

primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, e.g., chondrocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an

individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.*, Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response,

as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of a polypeptide of the invention (e.g., the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of

the invention and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, disorders characterized by aberrant expression or activity of the polypeptides of the invention include neurodegenerative disorders. In addition, the polypeptides of the invention can be used to promote nerve regeneration in the central and peripheral nervous system, to treat neurodegenerative disorders (e.g., Alzheimer's disease), memory loss, and neuronal damage (e.g., stroke), as well as other uses described herein. Moreover, the

HtrA-2 and TANGO 219 polypeptides of the invention can be used to modulate osteoblast proliferation and differentiation (e.g., to treat osteoporosis or arthritis).

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. For example, an antagonist of an HtrA-2 protein may be used to treat a proliferative disorder, e.g., cardiac hyperplasia, associated with aberrant HtrA-2 expression or activity. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder

characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect, e.g., in treatment of neurodegenerative disorders. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or upregulated and/or in which decreased activity is likely to have a beneficial effect, e.g., in treatment of an osteoblast proliferative disorder.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Deposit of Clones

Clones containing cDNA molecules encoding human HtrA-2 (clone EpT214) and human TANGO 219 (clone EpT219) were deposited with the American Type Culture Collection (Manassas, VA) on September 25, 1998 as Accession Number 98899, as part of a composite deposit representing a mixture of five strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (e.g., LB plates) supplemented with 100μg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

HtrA-2: 2.6 kb

TANGO 219: 1.3 kb

The identity of the strains can be inferred from the fragments liberated.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

International Application No: PCT/

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page __, lines _ of the description '

A. IDENTIFICATION OF DEPOSIT '

Further deposits are identified on an additional sheet '

Name of depositary institution '

American Type Culture Collection

Address of depositary institution (including postal code and country) '

10801 University Blvd.
Manassas, VA 20110-2209
US

Date of deposit ' September 25, 1998 Accession Number ' 98899

B. ADDITIONAL INDICATIONS ' (leave blank if not applicable). This information is continued on a separate attached sheet

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE ' (If the indications are not all designated States)

D. SEPARATE FURNISHING OF INDICATIONS ' (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later ' (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the International application when filed (to be checked by the receiving Office)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau "

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22, the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, 23 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, 23 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:3, 15, 20, 23 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899.

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, 23 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:2, 14, 19, 22, or a complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, 2, 13, 14, 18, 19, 21, 22 the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, 23, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number

98899.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 5, 20, or 23, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:3, 15, 20, or 23;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 14, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:2, 13, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:2, 14, 19, 22,or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:3, 15, 20, or 23.

10. The polypeptide of claim 8 further comprising heterologous amino acid

sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, or 23, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:3, 15, 20, or 23, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:3, 15, 20, or 23, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:3, 15, 20, or 23, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC as Accession Number 98899, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1 or 13, 18, 21, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of

claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for HtrA-2 or TANGO 219-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
GTCGACCCACGCGTCCGGGGAGCTGGTCCCTGCGCTCCCTGCGCCCTGGGGATGCCCCTGCCGCCCTGACGCCCGCCAG      79

CCTGAGCCACCGGCGCATGTGACCGCGCGTCCGCCCCAGTCCCATCCGTAGGCGCCCGGCGCCCGGCCCCGCAGCGGCC     158

                                                                    M   Q   A   R       4
TCGTTGTCCCCGCCGGCCCCCGCCCGGTCTCCCGCGCTGCCACCCGCCGCCGGCCCTGCCGCC ATG CAG GCG CGA     233

 A   L   L   L   A   A   L   A   A   L   A   L   A   R   E   P   P   A   A   P      24
GCG CTG CTC CTG GCC GCG TTG GCC GCG CTG GCG CTG GCC CGG GAG CCC CCT GCG GCG CCG     293

 C   P   A   R   C   D   V   S   R   C   P   S   P   R   C   P   G   G   Y   V      44
TGT CCC GCG CGC TGC GAC GTG TCG CGG TGT CCC AGC CCC CGC TGC CCC GGC GGC TAC GTG     353

 P   D   L   C   N   C   C   L   V   C   A   A   S   E   G   E   P   C   G   G      64
CCC GAC CTC TGC AAC TGC TGC CTG GTG TGC GCC GCC AGC GAG GGC GAG CCC TGT GGC GGC     413

 P   L   D   S   P   C   G   E   S   L   E   C   V   R   G   L   C   R   C   R      84
CCT CTG GAC TCG CCT TGC GGC GAG AGC CTG GAG TGC GTG CGC GGC CTA TGC CGC TGC CGC     473

 W   S   H   A   V   C   G   T   D   G   H   T   Y   A   N   V   C   A   L   Q     104
TGG TCG CAC GCC GTG TGT GGC ACC GAC GGG CAC ACC TAT GCC AAC GTG TGC GCG CTG CAG     533

 A   A   S   R   R   A   L   Q   L   S   G   T   P   V   R   Q   L   Q   K   G     124
GCG GCC AGC CGC CGC GCG CTG CAG CTC TCC GGG ACG CCC GTG CGC CAG CTG CAG AAG GGC     593

 A   C   P   L   G   L   H   Q   L   S   S   P   R   Y   K   F   N   F   I   A     144
GCC TGC CCG TTG GGT CTC CAC CAG CTG AGC AGC CCG CGC TAC AAG TTC AAC TTC ATT GCT     653
```

FIG.1A

```
 D   V   V   E   K   I   A   P   A   V   V   H   I   E   L   F   L   R   H   P    164
GAC GTG GTG GAG AAG ATC GCA CCA GCC GTG GTC CAC ATA GAG CTC TTC CTG AGA CAC CCG    713

 L   F   G   R   N   V   P   L   S   S   G   S   G   F   I   M   S   E   A   G    184
CTG TTT GGC CGC AAC GTG CCC CTG TCC AGC GGT TCT GGC TTC ATC ATG TCA GAG GCC GGC    773

 L   I   I   T   N   A   H   V   V   S   S   N   S   A   A   P   G   R   Q   Q    204
CTG ATC ATC ACC AAT GCC CAC GTG GTG TCC AGC AAC AGT GCT GCC CCG GGC AGG CAG CAG    833

 L   K   V   Q   L   Q   N   G   D   S   Y   E   A   T   I   K   D   I   D   K    224
CTC AAG GTG CAG CTA CAG AAT GGG GAC TCC TAT GAG GCC ACC ATC AAA GAC ATC GAC AAG    893

 K   S   D   I   A   T   I   K   I   H   P   K   K   K   L   P   V   L   L   L    244
AAG TCG GAC ATT GCC ACC ATC AAG ATC CAT CCC AAG AAA AAG CTC CCT GTG TTG TTG CTG    953

 G   H   S   A   D   L   R   P   G   E   F   V   V   A   I   G   S   P   F   A    264
GGT CAC TCG GCC GAC CTG CGG CCT GGG GAG TTT GTG GTG GCC ATC GGC AGT CCC TTC GCC   1013

 L   Q   N   T   V   T   T   G   I   V   S   T   A   Q   R   E   G   R   E   L    284
CTA CAG AAC ACA GTG ACA ACG GGC ATC GTC AGC ACT GCC CAG CGG GAG GGC AGG GAG CTG   1073

 G   L   R   D   S   D   M   D   Y   I   Q   T   D   A   I   I   N   Y   G   N    304
GGC CTC CGG GAC TCC GAC ATG GAC TAC ATC CAG ACG GAT GCC ATC ATC AAC TAC GGG AAC   1133

 S   G   G   P   L   V   N   L   D   G   E   V   I   G   I   N   T   L   K   V    324
TCC GGG GGA CCA CTG GTG AAC CTG GAT GGC GAG GTC ATT GGC ATC AAC ACG CTC AAG GTC   1193
```

FIG.1B

```
 T   A   G   I   S   F   A   I   P   S   D   R   I   T   R   F   L   T   E   F    344
ACG GCT GGC ATC TCC TTT GCC ATC CCC TCA GAC CGC ATC ACA CGG TTC CTC ACA GAG TTC   1253

 Q   D   K   Q   I   K   D   W   K   K   R   F   I   G   I   R   M   R   T   I    364
CAA GAC AAG CAG ATC AAA GAC TGG AAG AAG CGC TTC ATC GGC ATA CGG ATG CGG ACG ATC   1313

 T   P   S   L   V   D   E   L   K   A   S   N   P   D   F   P   E   V   S   S    384
ACA CCA AGC CTG GTG GAT GAG CTG AAG GCC AGC AAC CCG GAC TTC CCA GAG GTC AGC AGT   1373

 G   I   Y   V   Q   E   V   A   P   N   S   P   S   Q   R   G   G   I   Q   D    404
GGA ATT TAT GTG CAA GAG GTT GCG CCG AAT TCA CCT TCT CAG AGA GGC GGC ATC CAA GAT   1433

 G   D   I   I   V   K   V   N   G   R   P   L   V   D   S   S   E   L   Q   E    424
GGT GAC ATC ATC GTC AAG GTC AAC GGG CGT CCT CTA GTG GAC TCG AGT GAG CTG CAG GAG   1493

 A   V   L   T   E   S   P   L   L   L   E   V   R   R   G   N   D   D   L   L    444
GCC GTG CTG ACC GAG TCT CCT CTC CTA CTG GAG GTG CGG CGG GGG AAC GAC GAC CTC CTC   1553

 F   S   I   A   P   E   V   V   M   *    454
TTC AGC ATC GCA CCT GAG GTG GTC ATG TGA   1583

GGGGCGCATTCCTCCAGCGCCAAGCGTCAGAGCCTGCAGACAACGGAGGGCAGCGCCCCCCCGAGATCAGGACGAAGGA   1662

CCACCGTCGGTCCTCAGCAGGGCAGCAGCCTCCTCCTGGCTGTCCGGGGCAGAGCGGAGGCTGGGCTTGGCCAGGGGCC   1741

CGAATTTCCGCCTGGGGAGTGTTGGATCCACATCCCGGTGCCGGGGAGGGAAGCCCAACATCCCCTTGTACAGATGATC   1820
```

FIG.1C

CTGAAAGTCACTTCCAAGTTCTCCGGATATTCACAAAACTGCCTTCCATGGAGGTCCCCTCCTCTCCTAGCTTCCCGCC 1899

TCTGCCCCTGTGAACACCCATCTGCAGTATCCCCTGCTCCTGCCCCTCCTACTGCAGGTCTGGGCTGCCAAGCTTCTTC 1978

CCCCCTGACAAACGCCCACCTGACCTGAGGCCCCAGCTTCCCTCTGCCCTAGGACTTACCAAGCTGTAGGGCCAGGGCT 2057

GCTGCCTGCCAGCCTGGGGTCCCTGGAGGACAGGTCACATCTGATCCCTTTGGGGTGCGGGGGTGGGGTCCAGCCCAGA 2136

GCAGGCACTGAGTGAATGCCCCCTGGCTGCGGAGCTGAGCCCCGCCCTGCCATGAGGTTTTCCTCCCCAGGCAGGCAGG 2215

AGGCCGCGGGGAGCACGTGGAAAGTTGGCTGCTGCCTGGGGAAGCTTCTCCTCCCCAAGGCGGCCATGGGGCAGCCTGC 2294

AGAGGACAGTGGACGTGGAGCTGCGGGGTGTGAGGACTGAGCCGGCTTCCCCTTCCCACGCAGCTCTGGGATGCAGCAG 2373

CCGCTCGCATGGAAGTGCCGCCCAGAGGCATGCAGGCTGCTGGGCACCACCCCCTCATCCAGGGAACGAGTGTGTCTCA 2452

AGGGGCATTTGTGAGCTTTGCTGTAAATGGATTCCCAGTGTTGCTTGTACTGTATGTTTCTCTACTGTATGGAAAATAA 2531

AGTTTACAAGCACAAAAAAAAAAAAAAAAAAAAAAAGGGCGGCCGC 2577

FIG.1D

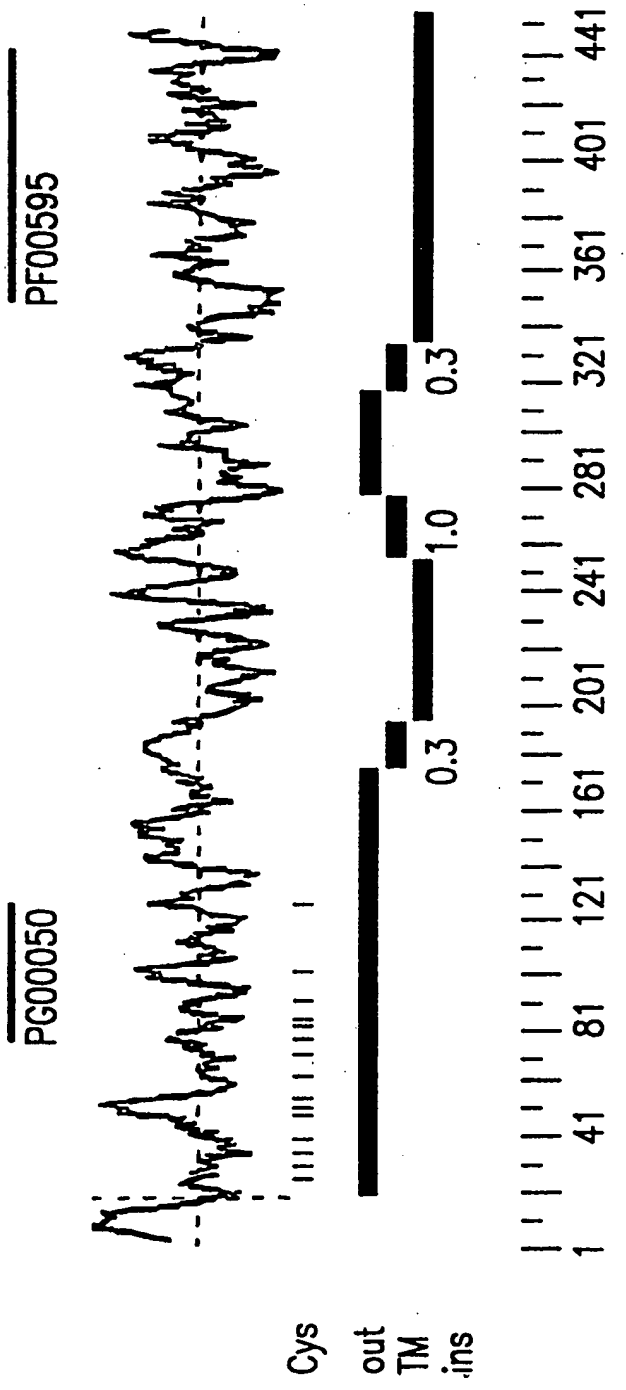
PG00050
PF00595
Cys
out
TM
ins
0.3
1.0
0.3
1
41
81
121
161
201
241
281
321
361
401
441

FIG.2A

MQARALLLAALAALALAREPPAAPCPARCDVSRCPS
PRCPGGYVPDLCNCCLVCAASEGEPCGGPLDSPCGE
SLECVRGLCRCRWSHAVCGTDGHTYANVCALQAAS
RRALQLSGTPVRQLQKGACPLGLHQLSSPRYKFNFIA
DVVEKIAPAVVHIELFLRHPLFGRNVPLSSGSGFIMSE
AGLIITNAHVVSSNSAAPGRQQLKVQLQNGDSYEATI
KDIDKKSDIATIKIHPKKKLPVLLLGHSADLRPGEFV
VAIGSPFALQNTVTTGIVSTAQREGRELGLRDSDMD
YIQTDAIINYGNSGGPLVNLDGEVIGINTLKVTAGISF
AIPSDRITRFLTEFQDKQIKDWKKRFGIRMRTITPSLV
DELKASNPDFPEVSSGIYVQEVAPNSPSQRGGIQDGD
IIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDL
LFSIAPEVVM

FIG.2B

```
          10                  20                  30        40        50
CGGGGTCCGCCGCCACCG-----CCG---CCGCCGCCAGAGTCGCCATGCAGATCCCGCG
 :  :.::  ::: .:::     : :   ::  :::       :::: :: .::: :: :
-GTCGACCCACGCGTCCGGGGAGCTGGTCCCTGCGCTCCCTGCGCCCTGGGGATGCCCCT
          10        20        30        40        50

      60          70          80          90          100
CGCCGCTCT-TCTCCCGC-TGC---TGCTGCTGCTGC---TGGCGGCGC--CCGCCTCGG
 ::::: :: .: ::::: .::    .:: .: :  ::   ::.: ::::  ::::: :.:
-GCCGCCCTGACGCCCGCCAGCCTGAGCCACCGGCGCATGTGACCGCGCGTCCGCCCCAG
60        70        80        90        100       110

      110                 120       130         140
-CGCAGCTGT-----CCCGG-GCC-GGCCGCTCGGCG-CCT--TTGGCC--GCCGGGT--
 : :: : ::      ::::: ::: :::: : :.::: :::  ::: ::  :::::
TCCCATCCGTAGGCGCCCGGCGCCCGGCCCCGCAGCGGCCTCGTTGTCCCCGCCGGCCCC
120       130       140       150       160       170

 150             160       170       180             190
-GCCCAGAC-----CGCTGCGAGCCGGCGCGCTGCCC-GCCGC--AGC---CG-GAGCAC
 ::::.:.:      :::::: : ::: ::: : :::: :::::  .::   :: :::::.:
CGCCCGGTCTCCCGCGCTGCCACCCGCCGC-CGGCCCTGCCGCCATGCAGGCGCGAGCGC
180       190       200       210       220       230

    200         210         220                 230       240
TGCGA-GGGCGGC-CGGGCC-CG-GGACGCGTGC--GG------CTGCTGCGAGGTGTG
:::    ::: ::   :::: ::  :.:::  ::  ::       :::: :::  ::::
TGCTCCTGGCCGCGTTGGCCGCGCTGGCGCTGGCCCGGGAGCCCCCTGCGGCGCCGTGTC
240       250       260       270       280       290

         250                 260         270         280
CGGCGCGC--C--CGAGG-GCG-----CCG---CGTGCGGCC--TGCAG--GAGGGCCCG
: ::::::  :  ::.:  :::     ::.   :  :: :::   ::.:  . : :::::
CCGCGCGCTGCGACGTGTCGCGGTGTCCCAGCCCCCGCTGCCCCGGCGGCTACGTGCCCG
300       310       320       330       340       350
```

FIG.3A

```
          290        300        310        320                   330
       T--GCGGCGAGGGGCTGCAGTGCGTGGTGC-CCTTCGGGG---TGCC----AGC----CT
       .   : ::.:  : ::::   : :::   : ::. ::.::   .:::    .::    ::
HtrA-2 ACCTCTGCAACTG-CTGCCTGGTGTGCGCCGCCAGCGAGGGCGAGCCCTGTGGCGGCCCT
       360        370        380        390        400        410

                  340        350             360        370
       C-GGCCACG--GTGCGGCG-GCGC-----GCGCAGGC-CGGCCTCTGTGTGTGC-GCCAG
       : :: :.::   ::::::: : ::     : ::. :: :::::: ::    ::: :: .:
HtrA-2 CTGGACTCGCCTTGCGGCGAGAGCCTGGAGTGCGTGCGCGGCCTATGCCGCTGCCGCTGG
        420        430        440        450        460        470

     380        390        400        410        420        430
       CAGCGAGCCGGTGTGCGGCAGCGACGCCAACACCTACGCCAACCTGTGCCAGCTGCGCGC
         ::. :::: :::: :::: :::::   ::::::: :::::: :::::  :::::. ::
HtrA-2 TCGCACGCCG-TGTGTGGCACCGACGGGCACACCTATGCCAACGTGTGCGCGCTGCAGGC
        480        490        500        510        520        530

     440        450        460        470          480        490
       CGCCAGCCGCCGCTCCGAGAGGCTGCACCGG-CCGCCGGT-CATCGTCCTGCAGCGCGGA
        :::::::::::: :  .: .::: :.:::::  :::: :: :. :. : ::::: . ::
HtrA-2 GGCCAGCCGCCGCGCGCTGCAGCT-CTCCGGGACGCCCGTGCGCCAGC-TGCAGAAGGGC
         540        550        560        570        580        590

       500        510          520        530        540        550
       GCCTGCGGCCAAGGG--CAGGAAGATCCCAACAGTTTGCGCCATAAATATAACTTTATCG
       ::::::  ::.. ::  :.  :  : :  :.:::   :::: : ::.:. ::::: :: :
HtrA-2 GCCTGC--CCGTTGGGTCTCCACCAGCTGAGCAGCCCGCGCTACAAGTTCAACTTCATTG
             600        610        620        630        640        650

         560        570        580        590        600        610
       CGGACGTGGTGGAGAAGATCGCCCCTGCCGTGGTTCATATCGAATTGTTTCGCAAGCTTC
       : :::::::::::::::::: ::.:::::::: :: :: ::. : :: :  :..:. :
HtrA-2 CTGACGTGGTGGAGAAGATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACC
             660        670        680        690        700        710
```

FIG.3B

```
          620        630        640        650        660        670
          CGTTTTCTAAACGAGAGGTGCCGGTGGCTAGTGGGTCTGGGTTTATTGTGTCGGAAGATG
          :: : : :.. :: .: ::::: :: : :: :: ::::: :: :: .::::.::.: :
HtrA-2    CGCTGTTTGGCCGCAACGTGCCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCG
               720        730        740        750        760        770

          680        690        700        710                720
          GACTGATCGTGACAAATGCCCACGTGGTGACCAACAA-------GCACCGGG--------
          : :::::::.: :: :::::::::::::::::.:::.:::        :: :::::
HtrA-2    GCCTGATCATCACCAATGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCAGGCAGC
               780        790        800        810        820        830

                    730        740        750        760        770
          ---TCAAAGTTGAGCTGAAGAACGGTGCCACTTACGAAGCCAAAATCAAGGATGTGGATG
             ::::.::  ::::. :::: :: : :.: :: ::.:::: :::::.:: .: :: .
HtrA-2    AGCTCAAGGTGCAGCTACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACA
               840        850        860        870        880        890

          780        790        800        810        820        830
          AGAAAGCAGACATCGCACTCATCAAAATTGACCACCAGGGCAAGCTGCCTGTCCTGCTGC
          ::::. :.:::::: ::   ::::::.::  : : : ::.. ::::: ::::: :: :::
HtrA-2    AGAAGTCGGACATTGCCACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGC
               900        910        920        930        940        950

          840        850        860        870        880        890
          TTGGCCGCTCCTCAGAGCTGCGGCCGGGAGAGTTCGTGGTCGCCATCGGAAGCCCGTTTT
          : :: :.:::  : :: :::::::: ::.:::::: ::::: :::::::: :: :: ::
HtrA-2    TGGGTCACTCGGCCGACCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCG
               960        970        980        990       1000       1010

          900        910        920        930        940        950
          CCCTTCAAAACACAGTCACCACCGGGATCGTGAGCACCACCCAGCGAGGCGGCAAAGAGC
          ::::.::.::::::::: :: :: :: ::::: ::::: .::::::::.:. ::::..::::
HtrA-2    CCCTACAGAACACAGTGACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCAGGGAGC
              1020       1030       1040       1050       1060       1070
```

FIG.3C

```
        960       970       980       990      1000      1010
        TGGGGCTCCGCAACTCAGACATGGACTACATCCAGACCGACGCCATCATCAACTATGGAA
        :::: ::::: .:::: :::::::::::::::::::: :: :::::::::::::: ::.:
HtrA-2 TGGGCCTCCGGGACTCCGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGA
            1080      1090      1100      1110      1120      1130

      1020      1030      1040      1050      1060      1070
        ACTCGGGAGGCCCGTTAGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTGAAAG
        :::: ::.:: ::. :.::.:::::::: :: ::.:: ::::: :: ::::: : ::.:
HtrA-2 ACTCCGGGGGACCACTGGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGG
            1140      1150      1160      1170      1180      1190

      1080      1090      1100      1110      1120      1130
        TGACAGCTGGAATCTCCTTTGCAATCCCATCTGATAAGATTAAAAAGTTCCTCACGGAGT
        : ::.:::::: :::::::::::: ::::: ::.::  . :: : : .::::::::::.::::
HtrA-2 TCACGGCTGGCATCTCCTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGT
            1200      1210      1220      1230      1240      1250

      1140      1150      1160      1170      1180      1190
        CCCATGACCGACAGGCCAAAGGAAAAGCCATCACCAAGAAGAAGTATATTGGTATCCGAA
         :::.::: ..::::. :::::.          :.  :::::: . :. :: :: :: ::.:
HtrA-2 TCCAAGACAAGCAGATCAAAGA---------CTGGAAGAAGCGCTTCATCGGCATACGGA
            1260      1270                1280      1290      1300

      1200      1210      1220      1230      1240      1250
        TGATGTCACTCACGTCCAGCAAAGCCAAAGAGCTGAAGGACCGGCA-CCGGGACTTCCCA
        ::  :.:. ::::. : ::: ..:  .:.::::::::::: ::.::: :: ::::::::::
HtrA-2 TGCGGACGATCACACCAAGCCTGGTGGATGAGCTGAAGG-CCAGCAACCCGGACTTCCCA
           1310      1320      1330      1340      1350      1360

        1260      1270      1280      1290      1300          1310
        GACGTGATCTCAGGAGCGTATATAATTGAAGTAATTCCTGATACCCCAGC----AGAAGC
        :: :: : :. .::::.  :::.:. ..:::.::..  :: .:::.: ::. :    :::.::
HtrA-2 GAGGTCAGCAGTGGAATTTATGTGCAAGAGGTTGCGCCGAATTCACCTTCTCAGAGAGGC
            1370      1380      1390      1400      1410      1420
```

FIG.3D

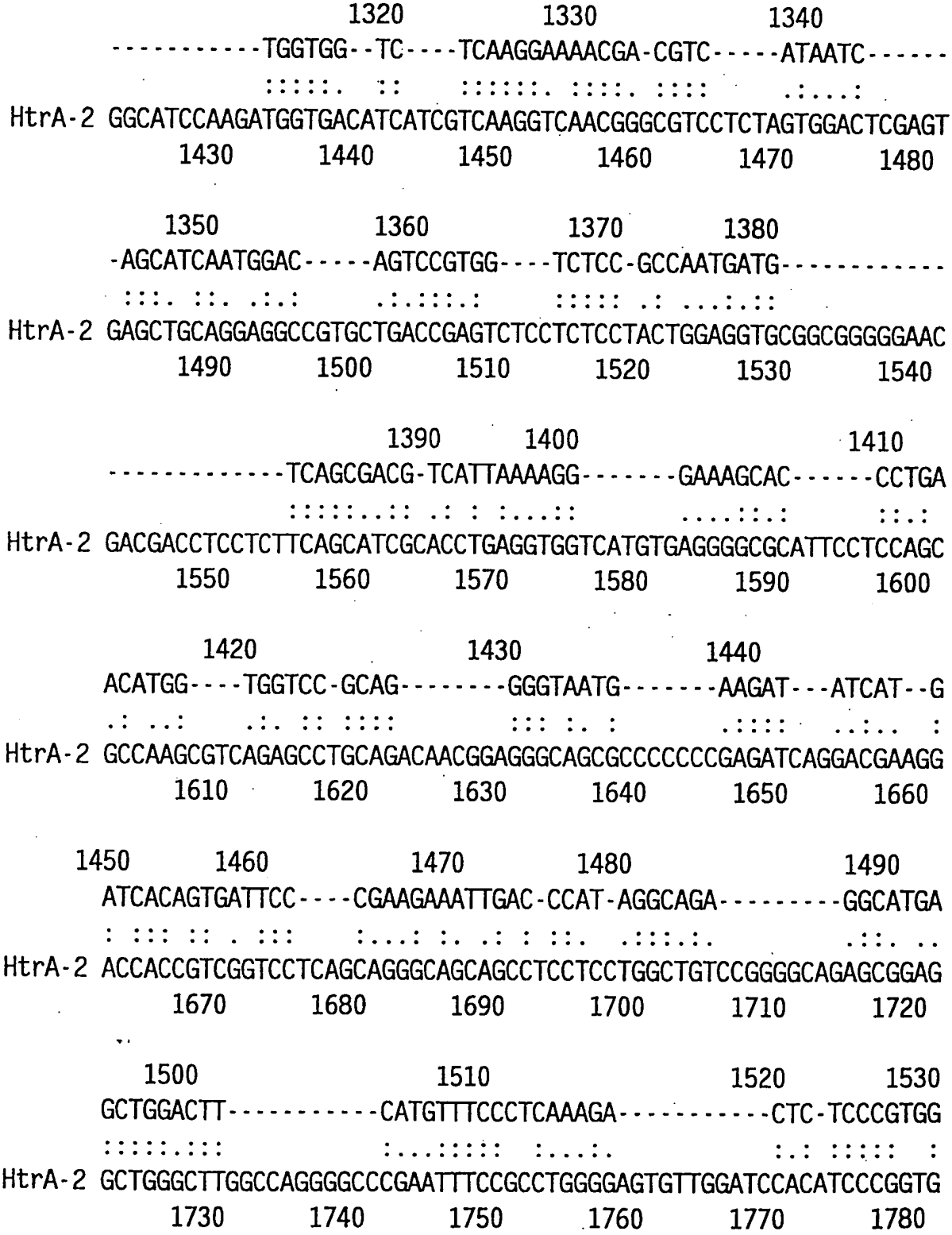
1320 1330 1340
----------TGGTGG--TC----TCAAGGAAAACGA-CGTC-----ATAATC------
:::::. :: ::::::. ::::. :::: .:...:
HtrA-2 GGCATCCAAGATGGTGACATCATCGTCAAGGTCAACGGGCGTCCTCTAGTGGACTCGAGT
1430 1440 1450 1460 1470 1480
1350 1360 1370 1380
-AGCATCAATGGAC-----AGTCCGTGG----TCTCC-GCCAATGATG----------
:::. ::. .:.: .:.::::.: ::::: .: ...:.::
HtrA-2 GAGCTGCAGGAGGCCGTGCTGACCGAGTCTCCTCTCCTACTGGAGGTGCGGCGGGGGAAC
1490 1500 1510 1520 1530 1540
1390 1400 1410
------------TCAGCGACG-TCATTAAAAGG-------GAAAGCAC------CCTGA
::::::.:: .: : :...:: ....::.: ::.:
HtrA-2 GACGACCTCCTCTTCAGCATCGCACCTGAGGTGGTCATGTGAGGGGCGCATTCCTCCAGC
1550 1560 1570 1580 1590 1600
1420 1430 1440
ACATGG----TGGTCC-GCAG--------GGGTAATG-------AAGAT---ATCAT--G
.: ..: .:. :: :::: ::: :. : .:::: ..:.. :
HtrA-2 GCCAAGCGTCAGAGCCTGCAGACAACGGAGGGCAGCGCCCCCCCGAGATCAGGACGAAGG
1610 1620 1630 1640 1650 1660
1450 1460 1470 1480 1490
ATCACAGTGATTCC----CGAAGAAATTGAC-CCAT-AGGCAGA---------GGCATGA
: ::: :: . ::: :...: :. .: : ::. .::::.:. .::. ..
HtrA-2 ACCACCGTCGGTCCTCAGCAGGGCAGCAGCCTCCTCCTGGCTGTCCGGGGCAGAGCGGAG
1670 1680 1690 1700 1710 1720
1500 1510 1520 1530
GCTGGACTT----------CATGTTTCCCTCAAAGA----------CTC-TCCCGTGG
:::::.::: :..:::::: :....:. :.: :::::
HtrA-2 GCTGGGCTTGGCCAGGGGCCCGAATTTCCGCCTGGGGAGTGTTGGATCCACATCCCGGTG
1730 1740 1750 1760 1770 1780

FIG.3E

```
          1540                1550        1560        1570
       ATGACGGATGAGGAC-----TC----TGGGCTGCTGGA---ATAGGACACT--CAAGA--
         :. :.. ::.: :      ::    :: .:.: ::..   ..:.:.:::: ::::.
HtrA-2 CCGGGGAGGGAAGCCCAACATCCCCTTGTACAGATGATCCTGAAAGTCACTTCCAAGTTC
          1790       1800       1810       1820       1830       1840

                  1580       1590                1600          1610
       ----------CTTTTGACTGCCATT-TTGTTTGT---------TCAGTGGA----GACTC
                :. ...:::::::.:  .:: . ::          ::  .: .    : :::
HtrA-2 TCCGGATATTCACAAAACTGCCTTCCATGGAGGTCCCCTCCTCTCCTAGCTTCCCGCCTC
          1850       1860       1870       1880       1890       1900

                           1620       1630       1640        1650
       ----CCTG---GC------CAACAGAATCC--TTCTTGATAGTTTGCAG--GCAAAACAA
           ::::    .:       :..:::.::::  : ::  .     : :..  :::...:..
HtrA-2 TGCCCCTGTGAACACCCATCTGCAGTATCCCCTGCTCCTGCCCCTCCTACTGCAGGTCTG
          1910       1920       1930       1940       1950       1960

          1660       1670       1680                1690
       A--TGTAATGTTGCAGATCCGCAGGCAGAAGCTCT----------GCCC----TTC--T
       .  ::  :.: : :.   :: :.:.::.: :: :.          ::::    :::  :
HtrA-2 GGCTGCCAAGCTTCTTCCCCCCTGACAAACGCCCACCTGACCTGAGGCCCCAGCTTCCCT
          1970       1980       1990       2000       2010       2020

        1700            1710            1720      1730
       GTATCCTATGT------ATGCAGT-------GTGCTTTTTCTTGCCAGCTTGGG---CCA
        :. :::: :.      :.:::.::       : :::  : : :::::::: ::::   ::.
HtrA-2 CTGCCCTAGGACTTACCAAGCTGTAGGGCCAGGGCTGCTGCCTGCCAGCCTGGGGTCCCT
          2030       2040       2050       2060       2070       2080

                  1740       1750                          1760
       -------------TTCTTG---CTTAGA----CAGT---------CAGC----ATTTGT
                    .::: .   ::::.:.    :.:          ::::    :  .:
HtrA-2 GGAGGACAGGTCACATCTGATCCCTTTGGGGTGCGGGGGTGGGGTCCAGCCCAGAGCAGG
          2090       2100       2110       2120       2130       2140
```

FIG.3F

```
                1770          1780          1790
      CTC----------CTCCTTT------AACTGAGTCAT---CATCTTAG------TCCAA
      :.:          : :::        :.:::::: :     :. : ..:      :::.
HtrA-2 CACTGAGTGAATGCCCCCTGGCTGCGGAGCTGAGCCCCGCCCTGCCATGAGGTTTTCCTC
           2150       2160       2170       2180       2190       2200

    1800        1810             1820       1830
      CTAATGCAGTC--GATAC-------AATGCGTAGATAGAAGA----AGCC----------
      :  : :::: :  :: .:       :. .:::.::.::..:.     .:::
HtrA-2 CCCAGGCAGGCAGGAGGCCGCGGGGAGCACGTGGAAAGTTGGCTGCTGCCTGGGGAAGCT
           2210       2220       2230       2240       2250       2260

                       1840        1850         1860       1870
      ---CCAC-------------GGG--AGCCAGGATGGGAC--TGGTCGTGTTTGTGC---
         ::.:             :::  :::::.: : .::::  :::.:::: .  :::
HtrA-2 TCTCCTCCCCAAGGCGGCCATGGGGCAGCCTGCAGAGGACAGTGGACGTGGAGCTGCGGG
           2270       2280       2290       2300       2310       2320

              1880        1890       1900       1910
      ---TTTTCTCCAAGTCAGCA--CCCAAAGGTCAATGCACAGAGA--C-----CC---CGG
         : .  .: .:: :.::.  ::::..   .:. .::.:.:.::  :     ::   ::
HtrA-2 GTGTGAGGACTGAGCCGGCTTCCCCTTCCCACGCAGCTCTGGGATGCAGCAGCCGCTCGC
           2330       2340       2350       2360       2370       2380

      1920                    1930                  1940       1950
      GTGGGTGAGC------------GCTGGCTTCT----------CAAACGGCCGAAGT---
      .:::.:.::              ::.::::: ::          :  .:. :::...:.
HtrA-2 ATGGAAGTGCCGCCCAGAGGCATGCAGGCTGCTGGGCACCACCCCCTCATCCAGGGAACG
           2390       2400       2410       2420       2430       2440

                    1960       1970        1980       1990
      ----TGCCTCTT------TTAGGAATCTCTT-TGGAATTGGGAGCACGATGA--CT----
         :: ::::..       ::.: .: :: :  :: ::.::::.. : :.::.  ::
HtrA-2 AGTGTGTCTCAAGGGGCATTTGTGAGCTTTGCTGTAAATGGATTCCCAGTGTTGCTTGTA
           2450       2460       2470       2480       2490       2500
```

FIG.3G

```
          2000                2010                2020
        CTGAGTTTGA----GCTATT---AAAGTAC----TTCTTACACATTG------------
        ::::.: : .    .::.:.   :::.::     :.:..::::...
HtrA-2 CTGTATGTTTCTCTACTGTATGGAAAATAAAGTTTACAAGCACAAAAAAAAAAAAAAAA
             2510       2520       2530       2540       2550       2560


        ---------------

HtrA-2 AAAAAAGGGCGGCCGC
             2570
```

FIG.3H

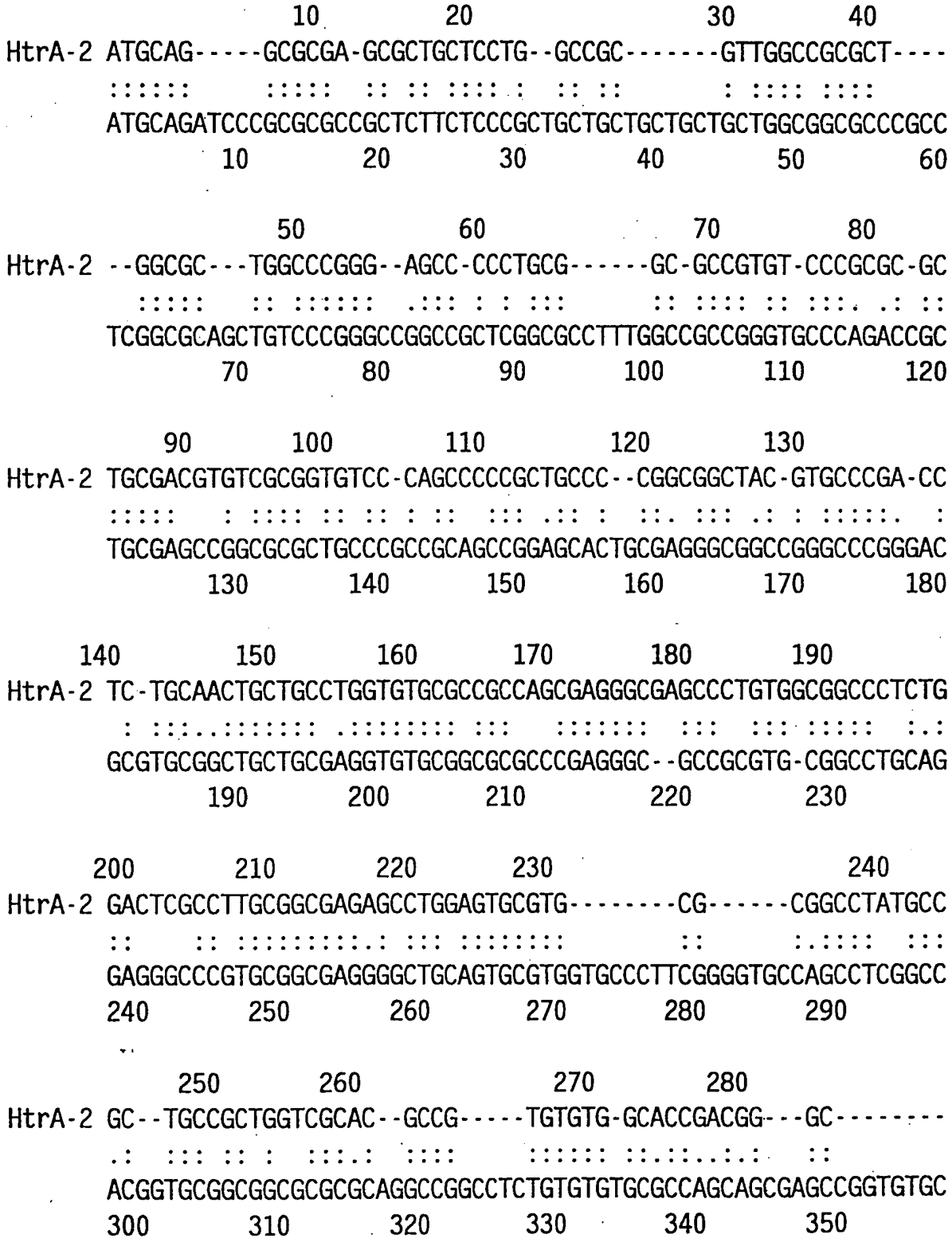
10 20 30 40
HtrA-2 ATGCAG-----GCGCGA-GCGCTGCTCCTG--GCCGC-------GTTGGCCGCGCT----
ATGCAGATCCCGCGCGCCGCTCTTCTCCCGCTGCTGCTGCTGCTGGCGGCGCCCGCC
10 20 30 40 50 60
50 60 70 80
HtrA-2 --GGCGC---TGGCCCGGG--AGCC-CCCTGCG------GC-GCCGTGT-CCCGCGC-GC
TCGGCGCAGCTGTCCCGGGCCGGCCGCTCGGCGCCTTTGGCCGCCGGGTGCCCAGACCGC
70 80 90 100 110 120
90 100 110 120 130
HtrA-2 TGCGACGTGTCGCGGTGTCC-CAGCCCCCGCTGCCC--CGGCGGCTAC-GTGCCCGA-CC
TGCGAGCCGGCGCGCTGCCCGCCGCAGCCGGAGCACTGCGAGGGCGGCCGGGCCCGGGAC
130 140 150 160 170 180
140 150 160 170 180 190
HtrA-2 TC-TGCAACTGCTGCCTGGTGTGCGCCGCCAGCGAGGGCGAGCCCTGTGGCGGCCCTCTG
GCGTGCGGCTGCTGCGAGGTGTGCGGCGCGCCCGAGGGC--GCCGCGTG-CGGCCTGCAG
190 200 210 220 230
200 210 220 230 240
HtrA-2 GACTCGCCTTGCGGCGAGAGCCTGGAGTGCGTG--------CG------CGGCCTATGCC
GAGGGCCCGTGCGGCGAGGGGCTGCAGTGCGTGGTGCCCTTCGGGGTGCCAGCCTCGGCC
240 250 260 270 280 290
250 260 270 280
HtrA-2 GC--TGCCGCTGGTCGCAC--GCCG-----TGTGTG-GCACCGACGG---GC--------
ACGGTGCGGCGGCGCGCGCAGGCCGGCCTCTGTGTGTGCGCCAGCAGCGAGCCGGTGTGC
300 310 320 330 340 350

FIG.4A

```
                          290        300        310        320        330
HtrA-2 ------------ACACCTATGCCAACGTGTGCGCGCTGCAGGCGGCCAGCCGCCGCGCG
                    ::::::: :::::: ::::: :::::. :: ::::::::::::
       GGCAGCGACGCCAACACCTACGCCAACCTGTGCCAGCTGCGCGCCGCCAGCCGCCGCTCC
       360        370        380        390        400        410

                340        350        360        370        380
HtrA-2 CTGCAGCT-CTCCGGGACGCCCGTGCGCCAGC-TGCAGAAGGGCGCCTGC--CCGTTGGG
        .: .::: :.:::::  :::: :: :. :. : ::::: . :: ::::::  ::.  .::
       GAGAGGCTGCACCGG-CCGCCGGT-CATCGTCCTGCAGCGCGGAGCCTGCGGCCA--AGG
       420        430        440        450        460        470

        390        400        410        420        430        440
HtrA-2 TCTCCACCAGCTGAGCAGCCCGCGCTACAAGTTCAACTTCATTGCTGACGTGGTGGAGAA
        :.  :  : :  :.:::   :::: : ::.:. ::::: :: :: ::::::::::::::
       GCAGGAAGATCCCAACAGTTTGCGCCATAAATATAACTTTATCGCGGACGTGGTGGAGAA
           480        490        500        510        520        530

        450        460        470        480        490        500
HtrA-2 GATCGCACCAGCCGTGGTCCACATAGAGCTCTTCCTGAGACACCCGCTGTTTGGCCGCAA
       :::::: ::.::::::::: :: :: ::. : :: :  :..:. ::: : : :.. :: .:
       GATCGCCCCTGCCGTGGTTCATATCGAATTGTTTCGCAAGCTTCCGTTTTCTAAACGAGA
           540        550        560        570        580        590

        510        520        530        540        550        560
HtrA-2 CGTGCCCCTGTCCAGCGGTTCTGGCTTCATCATGTCAGAGGCCGGCCTGATCATCACCAA
        :::::  :: : :: :: ::::: :: :: .::::.::.:  :: :::::::.: :: ::
       GGTGCCGGTGGCTAGTGGGTCTGGGTTTATTGTGTCGGAAGATGGACTGATCGTGACAAA
           600        610        620        630        640        650

        570        580        590        600        610        620
HtrA-2 TGCCCACGTGGTGTCCAGCAACAGTGCTGCCCCGGGCAGGCAGCAGCTCAAGGTGCAGCT
       :::::::::::::::.::.:::        :: :::::            ::::.::  ::::
       TGCCCACGTGGTGACCAACAA-------GCACCGGG-----------TCAAAGTTGAGCT
           660        670                680                690
```

FIG.4B

```
          630        640        650        660        670        680
HtrA-2 ACAGAATGGGGACTCCTATGAGGCCACCATCAAAGACATCGACAAGAAGTCGGACATTGC
       . :::: :: : :.: :: ::.::::  ::::::.:: .: :: .:::::. :.:::::: ::
       GAAGAACGGTGCCACTTACGAAGCCAAAATCAAGGATGTGGATGAGAAAGCAGACATCGC
         700        710        720        730        740        750

          690        700        710        720        730        740
HtrA-2 CACCATCAAGATCCATCCCAAGAAAAAGCTCCCTGTGTTGTTGCTGGGTCACTCGGCCGA
         :::::::.::  : : : ::.. ::::: :::::  :: :::: :: :.:::  : ::
       ACTCATCAAAATTGACCACCAGGGCAAGCTGCCTGTCCTGCTGCTTGGCCGCTCCTCAGA
         760        770        780        790        800        810

          750        760        770        780        790        800
HtrA-2 CCTGCGGCCTGGGGAGTTTGTGGTGGCCATCGGCAGTCCCTTCGCCCTACAGAACACAGT
         :::::::: ::.:::::: ::::: :::::::: :: :: ::  ::::.::.:::::::
       GCTGCGGCCGGGAGAGTTCGTGGTCGCCATCGGAAGCCCGTTTTCCCTTCAAAACACAGT
         820        830        840        850        860        870

          810        820        830        840        850        860
HtrA-2 GACAACGGGCATCGTCAGCACTGCCCAGCGGGAGGGCAGGGAGCTGGGCCTCCGGGACTC
         :: :: :: ::::: ::::: .:::::::.:. ::::.:::::::: ::::: .::::
       CACCACCGGGATCGTGAGCACCACCCAGCGAGGCGGCAAAGAGCTGGGGCTCCGCAACTC
         880        890        900        910        920        930

          870        880        890        900        910        920
HtrA-2 CGACATGGACTACATCCAGACGGATGCCATCATCAACTACGGGAACTCCGGGGGACCACT
         :::::::::::::::::::: :: :::::::::::::: ::.:::::: ::.:: ::. :
       AGACATGGACTACATCCAGACCGACGCCATCATCAACTATGGAAACTCGGGAGGCCCGTT
         940        950        960        970        980        990

          930        940        950        960        970        980
HtrA-2 GGTGAACCTGGATGGCGAGGTCATTGGCATCAACACGCTCAAGGTCACGGCTGGCATCTC
       .::.:::::::: :: ::.:: ::::: :: :::::  : ::.:: ::.:::::: :::::
       AGTAAACCTGGACGGTGAAGTGATTGGAATTAACACTTTGAAAGTGACAGCTGGAATCTC
         1000       1010       1020       1030       1040       1050
```

FIG.4C

```
        990       1000       1010       1020       1030       1040
HtrA-2 CTTTGCCATCCCCTCAGACCGCATCACACGGTTCCTCACAGAGTTCCAAGACAAGCAGAT
       :::::: ::::: ::.::  . :: : : .:::::::::.:::: :::.::: ..:::.
       CTTTGCAATCCCATCTGATAAGATTAAAAAGTTCCTCACGGAGTCCCATGACCGACAGGC
        1060       1070       1080       1090       1100       1110

       1050                 1060       1070       1080       1090
HtrA-2 CAAAGACTGG---------AAGAAGCGCTTCATCGGCATACGGATGCGGACGATCACACC
       :::::. ...           :::::: . :. :: :: :: ::.::: :.:. ::::. :
       CAAAGGAAAAGCCATCACCAAGAAGAAGTATATTGGTATCCGAATGATGTCACTCACGTC
        1120       1130       1140       1150       1160       1170

      1100       1110       1120       1130       1140       1150
HtrA-2 AAGCCTGGTGGATGAGCTGAAGG-CCAGCAACCCGGACTTCCCAGAGGTCAGCAGTGGAA
       ::: ..:  .:.:::::::::: ::.::: :: :::::::::::: :: : :. .:::.
       CAGCAAAGCCAAAGAGCTGAAGGACCGGCA-CCGGGACTTCCCAGACGTGATCTCAGGAG
        1180       1190       1200       1210       1220       1230

       1160       1170       1180       1190       1200       1210
HtrA-2 TTTATGTGCAAGAGGTTGCGCCGAATTCACCTTCTCAGAGAGGCGGCATCCAAGATGGTG
         :::.:. ..::.::.. :: .:::.: ::. :. :.. .:: ::  :: :.::... :
       CGTATATAATTGAAGTAATTCCTGATACCCCAGCAGAAGCTGGTGGTCTCAAGGAAAACG
         1240       1250       1260       1270       1280       1290

       1220       1230       1240       1250       1260       1270
HtrA-2 ACATCATCGTCAAGGTCAACGGGC-GTCCTCTAGTGGACTCGAGTGAG--CTGCAGGAGG
       ::.:::: .::::. .:::: ::.: ::::  :.::   : : :.:::   :.:: :. :
       ACGTCATAATCAGCATCAATGGACAGTCCG-TGGTCTCCGCCAATGATGTCAGC-GACGT
         1300       1310       1320       1330       1340       1350

       1280       1290       1300       1310       1320       1330
HtrA-2 CCGTGCTGACCGAGTCTCCTCTCCTACTGGAGGTGCGGCGGGGGAACGACGACCTCCTCT
       :  :. ...  .:. :.:: ::  .  :::.::: ::  :::: :: :: ::  :: : .
       CATTAAAAGGGAAAGCACC-CTGAACATGGTGGTCCGCAGGGGTAATGAAGATATCATGA
           1360       1370       1380       1390       1400       1410

                                  1340       1350
HtrA-2 TCAGCATCGCACCTGAGG---TGGTCATG
       :::  .: . .:: ::.:   : :.:  .
       TCACAGTGATTCCCGAAGAAATTGACCCA
            1420       1430       1440
```

```
                10        20        30        40        50        60
        MQIPRAALLPLLLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRARD
        :: .::       ::::: :.  :.:    .: :: ::.::. .::::..  .:.:: . :
HtrA-2 MQ-ARA-------LLLAALAALALAR----EPPAAPCPARCDVSRCPSP--RCPGGYVPD
                        10        20        30        40

                70        80        90       100       110
        ACGCCEVCGAPEGAACGLQ-EGPCGEGLQCVVPFGVPASATVRRRAQAGLCVCASSEPVC
         :.:: ::.:.::..:: . ..::::.:.::           :     ::: :  : .::
HtrA-2 LCNCCLVCAASEGEPCGGPLDSPCGESLECV----------R-----GLCRCRWSHAVC
          50        60        70                        80        90

      120       130       140       150       160       170
        GSDANTYANLCQLRAASRRSERLHRPPVIVLQRGACGQGQEDPNSLRHKYNFIADVVEKI
        :.:..::::.: :.::::::. .:   :: ::.::: :  . .: : :.:::::::::::
HtrA-2 GTDGHTYANVCALQAASRRALQLSGTPVRQLQKGACPLGLHQLSSPRYKFNFIADVVEKI
              100       110       120       130       140       150

      180       190       200       210       220               230
        APAVVHIELFRKLPFSKREVPVASGSGFIVSEDGLIVTNAHVVTN------KHRVKVELK
        :::::::::: . :.  :.::.:::::::.::.:::.:::::::..      .....:.:.
HtrA-2 APAVVHIELFLRHPLFGRNVPLSSGSGFIMSEAGLIITNAHVVSSNSAAPGRQQLKVQLQ
              160       170       180       190       200       210

         240       250       260       270       280       290
        NGATYEAKIKDVDEKADIALIKIDHQGKLPVLLLGRSSELRPGEFVVAIGSPFSLQNTVT
        ::..::: :::.: :.::: :::. . :::::::.:.:::::::::::::::::::::::
HtrA-2 NGDSYEATIKDIDKKSDIATIKIHPKKKLPVLLLGHSADLRPGEFVVAIGSPFALQNTVT
              220       230       240       250       260       270

         300       310       320       330       340       350
        TGIVSTTQRGGKELGLRNSDMDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVTAGISF
        :::::::.:: :.:::::.:::::::::::::::::::::::::::::::::::::::::
HtrA-2 TGIVSTAQREGRELGLRDSDMDYIQTDAIINYGNSGGPLVNLDGEVIGINTLKVTAGISF
              280       290       300       310       320       330
```

FIG.5A

```
          360        370        380        390        400        410
        AIPSDKIKKFLTESHDRQAKGKAITKKKYIGIRMMSLTSSKAKELKDRHRDFPDVISGAY
        ::::::.: .:::: .:.: :.    ::..:::::: ..:.: . :::. . ::::.: :: :
HtrA-2 AIPSDRITRFLTEFQDKQIKDW---KKRFIGIRMRTITPSLVDELKASNPDFPEVSSGIY
               340        350           360        370        380

          420        430        440        450        460        470
        IIEVIPDTPAEAGGLKENDVIISINGQSVVSANDVSDVIKRESTLNMVVRRGNEDIMITV
        . :: :..:.. ::....:.:. .::...:..... ...  :: : . :::::.:.....
HtrA-2 VQEVAPNSPSQRGGIQDGDIIVKVNGRPLVDSSELQEAVLTESPLLLEVRRGNDDLLFSI
        390        400        410        420        430        440

          480
        IPEEIDP
         :: .
HtrA-2 APEVVMN
        450
```

FIG.5B

```
AGGACAGTCTGCTTGTGGGCTCTGAAAGCTGGGGTGGGGCCAGAGCCTNAGCGNTTAATTTATTCABGHCHHCGYRDAC      79

                                M   V   S   F   R   K   I   F   I   L   Q   L   V      13
DCGTKKBTCCNACTTAAGGCAACAGC ATG GTT TCT TTC AGG AAA ATC TTC ATC TTA CAA CTT GTA      144

 G   L   V   L   T   Y   D   F   T   N   C   D   F   E   K   I   K   A   A   Y      33
GGG CTG GTG TTA ACT TAC GAC TTC ACT AAC TGT GAC TTT GAG AAG ATT AAA GCA GCC TAT      204

 L   S   T   I   S   K   D   L   I   T   Y   M   S   G   T   K   S   T   E   F      53
CTC AGT ACT ATT TCT AAA GAC CTG ATT ACA TAT ATG AGT GGG ACC AAA AGT ACC GAG TTC      264

 N   N   T   V   S   C   S   N   R   P   H   C   L   T   E   I   Q   S   L   T      73
AAC AAC ACC GTC TCT TGT AGC AAT CGG CCA CAT TGC CTT ACT GAA ATC CAG AGC CTA ACC      324

 F   N   P   T   A   G   C   A   S   L   A   K   E   M   F   A   M   K   T   K      93
TTC AAT CCC ACC GCC GGC TGC GCG TCG CTC GCC AAA GAA ATG TTC GCC ATG AAA ACT AAG      384

 A   A   L   A   I   W   C   P   G   Y   S   E   T   Q   I   N   A   T   Q   A     113
GCT GCC TTA GCT ATC TGG TGC CCA GGC TAT TCG GAA ACT CAG ATA AAT GCT ACT CAG GCA      444

 M   K   K   R   R   K   R   K   V   T   T   N   K   C   L   E   Q   V   S   Q     133
ATG AAG AAG AGG AGA AAA AGG AAA GTC ACA ACC AAT AAA TGT CTG GAA CAA GTG TCA CAA      504

 L   Q   G   L   W   R   R   F   N   R   P   L   L   K   Q   Q   *                 150
TTA CAA GGA TTG TGG CGT CGC TTC AAT CGA CCT TTA CTG AAA CAA CAG TAA                 555
```

FIG.6A

ACCATCTTTATTATGGTCATATTTCACAGCACCAAAATAAATCATCTTTATTAAGTAGATGAAACATTAACTCTAACTG 634

TGACAAAGAAGACCACAAATAGTTATCTTTTAATTACAGAAGAGTTTCTTAACTTACTTTTGTAAGTTTTTATTGTGTA 713

AGTTTATAATGCAGGGGAAGTACTACTCCTCAAATGTTGAGGGAAGCTTCCATAACATTGATGACTGGCTTCATGGCAG 792

NAATTCTCGGCTGTAGTTGCATAAGCATTGCTCAAGAGGAAAATCCAAAAGTGCAGCAGGAGAACTCTTTTCCCTGAAA 871

AAGGAAAAATATTGAACTCAATGATAGCACCTAAACTTACATTTAAAAGACAGACATTCCTTCTACATGTAATGACACT 950

TCTTGTGTTAAACTAAAAATTTACAAGAGAAGAAAGTGAAAGCAAATGGGGTTTCACAAATAGTTGTAAATATAGTGAA 1029

GCAATTTGAAATAATTTTCAAGCAAAGTATTGTGAAAGTATTCTAAGCCAAGTTTTAAATATTATCTAACAGACAAGAG 1108

TGGTATATACAAGTAGATCCTGAGAAGTACCTTTGTTACAGCTACTATAAATATACATATAAATTATAGAATCTACTTT 1187

AATTTATTTTGTGAACACTTTTGAAAATGTACATGTTCCTTTGTAATTGACACTATATATTTCTTAATAAAATAATTCT 1266

CA 1268

FIG.6B

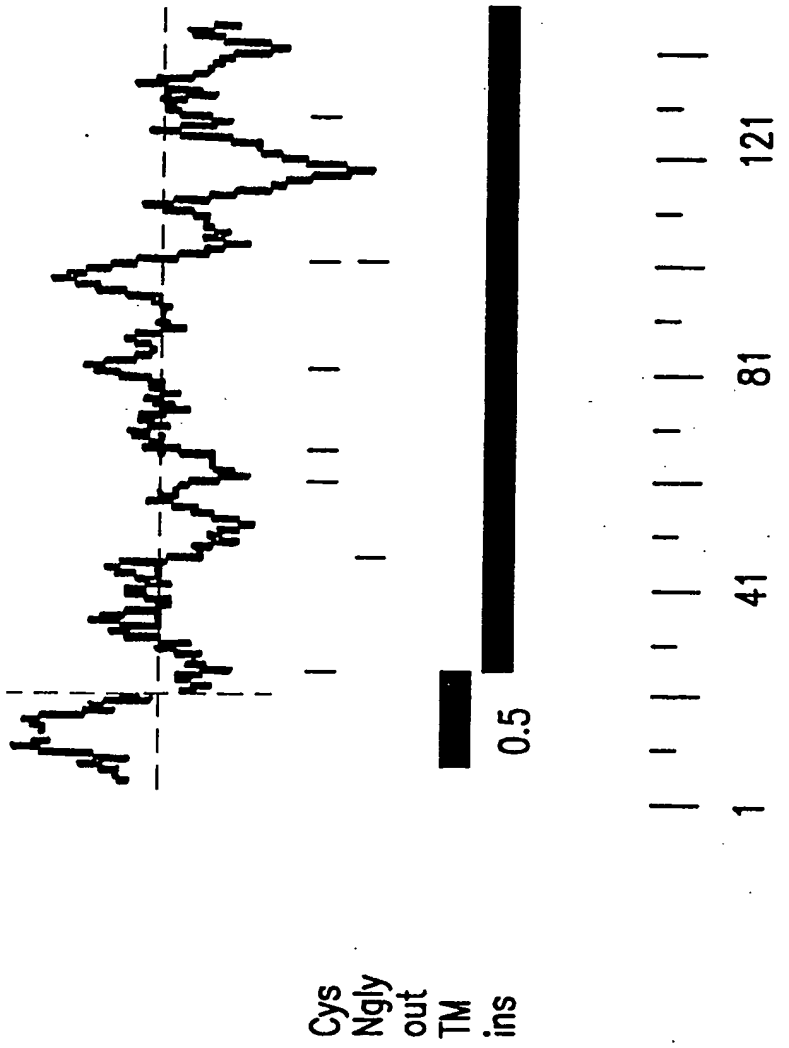

FIG.7A

MVSFRKIFILQLVGLVLTYDFTNCDFEKIKAAYLSTIS
KDLITYMSGTKSTEFNNTVSCSNRPHCLTEIQSLTFNP
TAGCASLAKEMFAMKTKAALAIWCPGYSETQINAT
QAMKKRRKRKVTTNKCLEQVSQLQGLWRRFNRPLL
KQQ

FIG.7B

```
GATTTTACCCCCCGGAAAGGAAGATGGAATTTTCTATGAAGGCCGACCTGGGAATGGGAAGATAAGAAAAGACCAGGAG     79

GGAGTTAAATAGGGAATGGGTTGGGGCGGCTTGGTAAATGTGCTGGGATTAGGCTGTTGCAGATAATGCAACAAGGCTT    158

GAAGGCTAACCTGGGGTGAGGCCGGGTTGGGGCCGGGCTGGGGGTGGGAGAAGTCTCACTGGCGGTTGATTGACAGTTT    237

                                    M  Q  A  R  A  L  Q  L  S  G  T  P                12
CTCCTTCCCCAGACTGGGGGTACCGCCACC ATG CAG GCG CGA GCG CTG CAG CTC TCC GGG ACG CCC     303

 V  R  Q  L  Q  K  G  A  C  P  L  G  L  H  Q  L  S  S  P  R                      32
GTG CGC CAG CTG CAG AAG GGC GCC TGC CCG TTG GGT CTC CAC CAG CTG AGC AGC CCG CGC    363

 Y  K  F  N  F  I  A  D  V  V  E  K  I  A  P  A  V  V  H  I                      52
TAC AAG TTC AAC TTC ATT GCT GAC GTG GTG GAG AAG ATC GCA CCA GCC GTG GTC CAC ATA    423

 E  L  F  L  R  H  P  L  F  G  R  N  V  P  L  S  S  G  S  G                      72
GAG CTC TTC CTG AGA CAC CCG CTG TTT GGC CGC AAC GTG CCC CTG TCC AGC GGT TCT GGC    483

 F  I  M  S  E  A  G  L  I  I  T  N  A  H  V  V  S  S  N  S                      92
TTC ATC ATG TCA GAG GCC GGC CTG ATC ATC ACC AAT GCC CAC GTG GTG TCC AGC AAC AGT    543

 A  A  P  G  R  Q  Q  L  K  V  Q  L  Q  N  G  D  S  Y  E  A                     112
GCT GCC CCG GGC AGG CAG CAG CTC AAG GTG CAG CTA CAG AAT GGG GAC TCC TAT GAG GCC    603

 T  I  K  D  I  D  K  K  S  D  I  A  T  I  K  I  H  P  K  K                     132
ACC ATC AAA GAC ATC GAC AAG AAG TCG GAC ATT GCC ACC ATC AAG ATC CAT CCC AAG AAA    663
```

FIG.8A

```
K  L  P  V  L  L  L  G  H  S  A  D  L  R  P  G  E  F  V  V   152
AAG CTC CCT GTG TTG TTG CTG GGT CAC TCG GCC GAC CTG CGG CCT GGG GAG TTT GTG GTG   723

A  I  G  S  P  F  A  L  Q  N  T  V  T  T  G  I  V  S  T  A   172
GCC ATC GGC AGT CCC TTC GCC CTA CAG AAC ACA GTG ACA ACG GGC ATC GTC AGC ACT GCC   783

Q  R  E  G  R  E  L  G  L  R  D  S  D  M  D  Y  I  Q  T  D   192
CAG CGG GAG GGC AGG GAG CTG GGC CTC CGG GAC TCC GAC ATG GAC TAC ATC CAG ACG GAT   843

A  I  I  N  Y  G  N  S  G  G  P  L  V  N  L  D  G  E  V  I   212
GCC ATC ATC AAC TAC GGG AAC TCC GGG GGA CCA CTG GTG AAC CTG GAT GGC GAG GTC ATT   903

G  I  N  T  L  K  V  T  A  G  I  S  F  A  I  P  S  D  R  I   232
GGC ATC AAC ACG CTC AAG GTC ACG GCT GGC ATC TCC TTT GCC ATC CCC TCA GAC CGC ATC   963

T  R  F  L  T  E  F  Q  D  K  Q  I  K  D  W  K  K  R  F  I   252
ACA CGG TTC CTC ACA GAG TTC CAA GAC AAG CAG ATC AAA GAC TGG AAG AAG CGC TTC ATC  1023

G  I  R  M  R  T  I  T  P  S  L  V  D  E  L  K  A  S  N  P   272
GGC ATA CGG ATG CGG ACG ATC ACA CCA AGC CTG GTG GAT GAG CTG AAG GCC AGC AAC CCG  1083

D  F  P  E  V  S  S  G  I  Y  V  Q  E  V  A  P  N  S  P  S   292
GAC TTC CCA GAG GTC AGC AGT GGA ATT TAT GTG CAA GAG GTT GCG CCG AAT TCA CCT TCT  1143

Q  R  G  G  I  Q  D  G  D  I  I  V  K  V  N  G  R  P  L  V   312
CAG AGA GGC GGC ATC CAA GAT GGT GAC ATC ATC GTC AAG GTC AAC GGG CGT CCT CTA GTG  1203
```

FIG.8B

```
 D   S   S   E   L   Q   E   A   V   L   T   E   S   P   L   L   L   E   V   R    332
GAC TCG AGT GAG CTG CAG GAG GCC GTG CTG ACC GAG TCT CCT CTC CTA CTG GAG GTG CGG   1263

 R   G   N   D   D   L   L   F   S   I   A   P   E   V   V   M   *                349
CGG GGG AAC GAC GAC CTC CTC TTC AGC ATC GCA CCT GAG GTG GTC ATG TGA              1314

GGGGCGCATTTCTCCAGCGCCAAGCGTCATCGATTACTCGAGCCATGCGACCCCACGCCACCCCGTGCCTCCTGCCTCC  1393

GCGCAGCCTGCAGCGGGAGACCCTGTCCCCGCCCAGCCGTCCTCTGGGGTGGACCTAGTTTAATAAAGATTACCAAGTT  1472

TCACGCATCTGCTGGCCTCCCCTGTGATTTCCTCTAAGCCCCACCTTAGTTTCTCTTTCTGCCCATACTGGGCACACAA  1551

TTCTCAGCCCCT                                                                     1563
```

FIG.8C

```
    L   I   P   D   T   F   L   V   S   F   P   P   I   P   I   P   F   P   V     19
  T CTG ATA CCA GAC ACC TTC CTG GTG TCT TTC CCT CCT ATC CCC ATC CCC TTC CCT GTC     58

  P   F   I   Q   F   L   I   S   G   G   F   N   L   L   S   L   S   N   C   A     39
CCT TTC ATT CAA TTT TTA ATA TCT GGC GGT TTT AAT CTT CTC TCT CTC TCG AAC TGT GCC    118

  A   C   E   Q   P   A   C   L   L   K   I   E   Y   Y   T   L   N   P   I   P     59
GCT TGT GAG CAG CCA GCT TGT CTC CTG AAA ATC GAG TAT TAT ACT CTC AAT CCT ATC CCT    178

  G   C   P   S   L   P   D   K   T   F   A   R   R   T   R   E   A   L   N   D     79
GGC TGC CCT TCA CTC CCC GAC AAA ACA TTT GCC CGG AGA ACA AGA GAA GCC CTC AAT GAC    238

  H   C   P   G   Y   P   E   T   E   R   N   D   G   T   Q   E   M   A   Q   E     99
CAC TGC CCA GGC TAC CCT GAA ACT GAG AGA AAT GAC GGT ACT CAG GAA ATG GCA CAA GAA    298

  V   Q   N   I   C   L   N   Q   T   S   Q   I   L   R   L   W   Y   S   F   M    119
GTC CAA AAC ATC TGT CTG AAT CAA ACC TCA CAA ATT CTA AGA TTG TGG TAT TCC TTC ATG    358

  Q   S   P   E   *                                                               124
CAA TCT CCA GAA TAA                                                             373

AATTAGCTTTCAGCTTCTGCTATG                                                        397
```

FIG.9

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☑ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.